UNIVERSITAT DE BARCELONA

Scientific explanation in biology. Beyond mechanistic explanation

Roger Deulofeu Batllori

UNIVERSITAT DE BARCELONA

SCIENTIFIC EXPLANATION IN BIOLOGY. BEYOND MECHANISTIC EXPLANATION

ROGER DEULOFEU BATLLORI

PhD Dissertation

Supervisor:
JOSE A. DIEZ CALZADA

University of Barcelona

Faculty of Philosophy

PhD Programme: Cognitive Science and Language

(EEES:HDK08 Ciència Cognitiva i Llenguatge)

Line of Investigation: Analytic philosophy

(101133 Filosofia analítica)

December 2019

AGRAIMENTS

Aquesta tesi no hauria estat possible sense l'ajuda constant de varies persones. Agraeixo al José A. Díez, director de la tesi, per la seva constància i suport, i perquè sempre m'ha transmès confiança en que aquest projecte tiraria endavant. A la gent de Logos, i en concret a la gent del grup de filosofia de la ciència, el Carl, la Romina, l'Albert, el Johannes, l'Alfonso, la Pilar… perquè conversant amb ells tot analitzant articles al grups de lectura de ciència és com més n'he après i com més he gaudit fent filosofia. També als doctorands del GRG i de la oficina 4090, amb qui he tingut la sort de discutir algunes de les meves recerques, i en especial al Johannes, que ens ha passat feedback valuós dels articles en les seves versions inicials. A l'Albert Solé, perquè sense ell no hauria arribat a Logos, i probablement no hagués obtingut la FPI ni l'empenta necessària per a llançar-me a un projecte d'aquestes dimensions. Al Michele Palmira, que ha fet que els dies d'investigador a la facultat fossin mes interessants i a l'hora simpàtics, i pels seus suport, consells i comentaris a les últimes versions de la tesi. La Roser, la meva parella, sempre ha estat al meu costat, tant en els bons moments com en els difícils, sobretot els de l'estada a Toronto. Sense ella, els seus ànims i la seva confiança en mi això mai hagués tirat endavant. Al meu pare, que sempre ha estat insistint, preguntant i aconsellant, i a la meva mare, pels ànims i per posar sempre bona cara als moments difícils. Voldria agrair molt especialment al Javier Suárez, amb qui he estat compartint tots aquests anys d'investigador a la facultat. Ell em va contagiar la passió per la filosofia de la biologia, i treballant al seu costat no només he après un munt de coses i he gaudit com mai de les discussions que hem tingut (i no només de filosofia) sinó que puc dir que el filòsof en el que m'he convertit és en gran part gràcies a ell.

Finalment, aquesta recerca no hagués estat possible sense el suport econòmic de la beca FPI (BES-2013-063239) que em va concedir el 'Ministerio de Educación, Cultura y Deporte', així com als projectes de recerca *Representación y explicación en la ciencia: análisis monistas y pluralistes* (FFI2012-37354) y *Leyes, explicación y realismo en ciencias físicas y biomédicas* (FFI2016-76799-P). A ells també els hi ho agraeixo.

CONTENTS

I. ABSTRACT

Understanding how scientists explain has been one of the major goals of the philosophy of science. Given that explaining is one of the most important tasks that scientists aim at and given the high specialization that currently affects all scientific disciplines, we encounter what might at first glance appear to us as many different types of explanations and very different ways of explaining natural phenomena. This suggests a pluralist picture regarding scientific explanation, particularly in biology, namely the existence of different accounts of explanation that do not share an interesting common core. However, the main goal of the traditional analysis of scientific explanation was to elaborate a monist theory of explanation according to which all scientific explanations share a common core that makes them what they are - i.e. that they can be identified by a commonly shared set of necessary and jointly sufficient conditions. The monist accounts mainly draw on examples from physics to illustrate how this is supposed to work, leaving examples from the special science, like biology, aside.

In the last twenty years, nonetheless, the rise of the New Mechanism philosophy, with its notion of mechanistic explanation, has become the dominant and widely accepted account among the philosophers of science to analyze scientific explanation in biology, challenging the pluralist view. The New mechanist account of scientific explanation is essentially monist since their defenders claim that mechanisms are all what really matters to explanation. According to mechanistic explanation, in order to explain a biological phenomenon, we have to discover the mechanism that is responsible for it. Further, we have to decompose this mechanism in order to identify its component parts and identify the causal story that connects the components with the phenomenon. Mechanistic explanations are thus considered causal explanations.
The New Mechanism philosophy has arguably been very successful in analyzing how explanation works in a huge diversity of models in biology, suggesting that their account of mechanistic explanation is the only legitimate of in biology. Furthermore, New Mechanism philosophy provides a new framework that contributed to tackle traditional problems of the philosophy of science related to notions such as laws of nature, function, causation, etc.

Although mechanistic explanation has proved very successful in analyzing the explanatory force of many biological models, its scope in biology is still under discussion. In the last few years, there has been voices limiting the extension of this account. On the one hand, there has been philosophers claiming that in some biological models, mathematics plays not only a representational role but an explanatory role, suggesting that those models provide explanations that rather than identifying a mechanism with its components and causal story, identify mathematical properties that are explanatory of some phenomenon. They claim that in those explanations, the system under analysis has a mathematical structure whose mathematical properties are explanatory of a particular range of *explananda*. On the other hand, and despite the claim widely accepted that there are no laws in biology, some philosophers claim we can still consider that some biological models explain by appeal to laws of nature, suggesting covering law accounts of scientific explanation.

The present thesis dissertation is a contribution to the aforementioned debate. It provides examples of biological models whose explanatory power does not lie in its identification of mechanisms with its parts and causal story, even if the models look somehow mechanistic. I claim they provide non-mechanistic (and non-causal) explanations, in so far as the models, even if they could identify a mechanism, do not explain by pinpointing information about its causal story.

1. INTRODUCTION. SCIENTIFIC EXPLANATION IN BIOLOGY. BEYOND THE SCOPE OF MECHANISTIC EXPLANATION

The aim of this introduction is to offer the philosophical debates in which the three papers that conform this thesis dissertation engage in. §1.1 provides a general discussion about the main goals of the traditional accounts of scientific explanation and the role they play in recent debates about explanation in biology. §1.2 introduces the notion of mechanistic explanation in detail. §1.3 presents some problems raised against mechanistic explanation following two different strategies related to, mathematical explanation on the one hand and a neo-Hempelian account of explanation on the other. §1.4 discusses the concept of explanatory pluralism in biology. §1.5 gives final remarks.

1.1. From general scientific explanation towards explanation in biology

Since the appearance of the Deductive Nomological model (DN, hereafter) (Hempel Oppenheim 1948; Hempel 1962, 1965) the philosophical interest in characterizing scientific explanations has been articulated around three main questions (Braillard & Malaterre 2015, p.3):

1) Are there unique features of scientific explanation, in terms of necessary and sufficient conditions?
2) Is causation (or should be) a primitive notion for scientific explanation?
3) Which role does context play in scientific explanation?

In this section, I will briefly introduce the answers to questions 1 and 2 that have been given by some of the most influential philosophical accounts of scientific explanation[1,2]. These accounts have a monist flavour, and universalist aspirations, as they were believed to apply to all instances of scientific explanation. As this brief detour will show, besides the problems each monist account faces, contemporary debates in biological explanation seem to be at odds with the claim that there is just *one philosophical account* of scientific explanation that normatively tells what a genuine scientific explanation is. This is why the traditional monistic picture of scientific explanation has recently been systematically replaced by a pluralist view of the phenomenon of scientifically explaining (Braillard & Malaterre 2015; Moreno & Suárez (submitted)).

Let us begin by addressing question 1. The aim of this question is to find a way to draw a line, in terms of necessary and sufficient conditions, between *genuine* scientific explanations and apparent one. This goal was a departure point for several philosophers of science that started to think about explanation (Hempel & Oppenheim 1948). In "Four decades of scientific explanation" (1989), Wesley Salmon provides an extensive summary of all the attempts to provide necessary and sufficient conditions for a scientific explanation. Unfortunately, Salmon's paper (1989) shows that despite four decades (1950-1990) of hard and dedicated work, this goal was far from being achieved by any account.

---

[1] Questions 1 and 2 are directly related to the topic of this dissertation. Question 3 is somehow orthogonal to the debates engaged in this dissertation, this is why I will not speak about it.

[2] Notice that the presentation of the traditional accounts of scientific explanation in this section will be rather schematic. My aim is not to fully analyze how they work but to depict a story about the attempt to construe monist accounts of scientific explanation and how it failed.

Carl Hempel's DN model is the first and most discussed account of scientific explanation that pursue such an aim. According to Hempel (1962, 1965; Hempel & Oppenheim 1948), a scientific explanation must take the form of a deductive argument. The DN model has two main epistemic elements. The *explanandum*, which is the phenomenon to explain, must be the conclusion of the argument; the *explanans*, which is the set of premises that deductively entail the *explanandum*. According to the DN model, the *explanans* must contain at least one law of nature (law-like statements to be more accurate), and some of the premises must have empirical content. The appeal to laws of nature in the *explanans* comes from the empiricist attempt to avoid the appeal to the notion of cause, for causation was thought to be more metaphysically loaded than laws[3]. In particular, the fundamental worry for Hempel (1962, 1965) and other empiricists (Nagel 1961; Carnap 1966) was to draw a line between nomological generalizations (i.e., law of nature) and accidental generalizations, so that they could distinguish genuine scientific explanations from pseudo-explanations[4].

Moreover, as many cases involving statistical generalizations were not covered by the DN model, Hempel elaborated a complementary model, the Inductive Statistical model (IS, hereafter) of scientific explanation. The IS model is analogous to the DN model with the sole different that the premises of the argument include a statistic or probabilistic generalization in their *explanans*, rather than a universal generalization (Hempel 1965). According to this model, explanations work by showing how the *explanans* confers a high degree of probability to the *explanandum*.

DN model and IS model received several counterexamples:

Symmetry problem: The flagpole. The flagpole case illustrates an scenario in which the explanation of the length of the shadow of a flagpole projected in the floor appeals to the height of the pole, the angles of the sunlight incidence and some geometric laws; it turns out, though, that you can also "obtain" the height of the pole by appealing to the length of the shadow, the angle of the sunlight incidence and the same geometric laws. Interestingly, both deductions satisfy the DN model conditions, although no one would intuitively accept that the second is a genuine explanation (cf. Van Fraassen 1980, pp. 130-134). The problem here lies in the fact that explanations are in most cases asymmetric, and the DN model works symmetrically in several cases.

Irrelevance: The pregnant male. The "birth control pills" case illustrates an explanation of why a specific male did not get pregnant, by appealing to the fact that it took birth control pills. Here, the fact that a given male did not get pregnant is inferred from the non-accidental generalization "all males *who take birth control pills* fail to get pregnant" and the premise stating that the male took birth control pills, fulfilling the inference the structure of the DN model. Thus, here we have

[3] However, as posterior literature has shown (Dretske 1977; Cartwright 1980; Armstrong 1983; Cohen & Callender 2009; Carrol 2016), philosophers of science are far from a consensus regarding the notion of a law of nature.

[4] Neither Hempel, nor Nagel or other empiricist managed to demarcate in a compelling way accidental from non-accidental generalization without appealing to loaded metaphysical notions such as nomological necessity (Nagel 1961; Carnap 1966).

the case of an explanations that fulfils the conditions of DN model but that fails to be a good explanation for it appeals to irrelevant facts as explanatory for the *explanandum*

Common cause: The barometer. The barometer-storm case provides the prediction/inference of a storm by appealing to a sudden drop in the measure of a barometer and the law that whenever there is a sudden drop in the measure of a barometer a storm happens in the surrounding area. This inference fulfils the pattern of DN model but no one would accept that the sudden drop in the barometer explains the storm: what explains the storm is a common cause, namely the sudden drop in atmospheric pressure.

Temporal asymmetry: The case of the eclipse. We can infer a solar eclipse from the positions, velocities and masses of the Moon, the Sun and the Earth previous to the eclipse and the laws of planetary motion. However, we can as well infer the eclipse from the positions, velocities and masses of the Sun, Moon and Earth after the eclipse and the same laws of planetary motion. The first and not the second would be considered an explanation of the eclipse, despite the fact that both fulfil the requirements of the DN model.

Challenges to the IS model: The case of Mayor with paresis. The case of paresis illustrates that some explanations may rely in *explanans* that, despite providing low probability to the *explanandum*, are considered explanatory for its relevance. In this case, the fact that the mayor got paresis is explained by the premise that the mayor had syphilis and she was not treated with penicillin, plus the probabilistic law stating that around 25% of patients that had syphilis without getting a treatment develop paresis. The inductive argument is fallacious, for the premises confer few probabilities to the conclusion, although the explanation is considered valid[5].

Thus, we can conclude after the counter examples that Hempel's models are over-permissive and too restrictive. The paresis case shows that IS model does not provide necessary conditions for statistical explanations (non-necessity problems), while the others show that DN model does not provide sufficient conditions for deterministic explanations (non-sufficiency problems).

The existence of these counterexamples and other worries, such as the difficulties to find a compelling empiricist analysis of laws of nature, makes DN model and IS model controverted models of explanation. Contemporary discussions about explanation in biology do not take these models in much consideration, turning the focus on causal explanations. Regardless of the actual value that the Hempelian model has on its own, DN and IS models are also important because the philosophical debates around scientific explanation were always motivated as a criticism or as a support to them. It is important as well for it is the first philosophical attempt to account for the phenomenon of scientific explanation with universalist aspirations. Namely, Hempel thought that all instances of deterministic/statistical explanations in science should fulfil all the criteria described above. Despite the criticisms, there is yet one main feature shared by the DN model and the IS model that is still considered useful and worth saving by some contemporary accounts of scientific explanation (Díez 2014; Alleva et al. 2017; Deulofeu & Suárez 2018), namely, explaining a phenomenon is to show how its occurrence is to be expected on the basis of lawful generalizations (more details in section 1.3.2).

---

[5] See Salmon (1971, 1989), Díez & Moulines (2008, ch.7) and Woodward (2017) for a detailed presentation of the counterexamples to DN and IS models, as well as its source.

Drawing upon question 2, which inquiries about the role of causation in scientific explanation, some philosophers argued that the DN model and the IS model were unsatisfactory precisely because they were avoiding the appeal to causal relations (Scriven 1962). Because of that, Salmon (1971, 1984, ch. 2) displaced the notion of natural law as the cornerstone of scientific explanation and elaborated a philosophical model based on the notion of *cause*, in particular, on the notion of *causal relevance*. The Statistical Relevant Model (SR, hereafter) (Salmon 1971, 1984, ch. 2) was trying to capture the causal relevant factors for the explanation of a phenomenon that were left aside by Hempel's DN model and IS model. Woodward summarizes the SR model as follows: "Given some class or population *A*, an attribute *C* will be *statistically relevant* to another attribute *B* if and only if $P(B \mid A.C) \neq P(B \mid A)$- that is, if and only if the probability of *B* conditional on *A* and *C* is different from the probability of *B* conditional on *A* alone" (2017: 24). The main idea of the SR model is that statistical relevant properties are explanatory while statistically irrelevant are not, or in other words, "the notion of a property making a difference for an explanandum is unpacked in terms of statistical relevant relationships" (Woodward 2017, p. 24). Moreover, Salmon claims that in cases of explanations involving statistical generalizations, the relation between *explanans* and *explanandum* is not the one of high probability, as advanced by IS model, but rather of statistical relevance.

Salmon's model managed to overcome the cases of explanatory irrelevance, of explanatory asymmetries and explanations with low probability. For instance, with respect to the case of *irrelevance*, it is clear that for males, the fact of having birth control pills, is statistically irrelevant for the fact of not getting pregnant. Regarding problems of symmetries (flagpole), the asymmetry of causal relations seems to block the problems this counterexample illustrates. The flagpole is the cause of the shadow and not the other way around. With respect to the low probability events, the case of the Mayor with paresis, Salmon's SR model shows that even if the *explanans* confer low probability to the explanandum, the explanation is still valid for it raises the probabilities of the phenomenon, even if the raise is small.

However, SR model has its own problems. Woodward (2017) claims that if we look for the motivation roots of such a model, we can end up with two main ideas: (i) explanation must cite causal relationships and (ii) causal relationships are captured by statistical relevance relationships (2017). Even if (i) could be accepted, others have shown (Cartwright 1979) that (ii) is false, for causal relations are underdetermined by statistical relations. Thus, that causal relations can be captured by statistical relevant relations is at least highly controversial.

Because of these problems, Salmon (1984) elaborated another philosophical model of scientific explanation called the Causal Mechanical Model (CM, hereafter) which is the departure point of the New Mechanism philosophy (discussed later, section 1.2), although CM model focuses more on etiological rather than in constitutive causal explanations[6] (Salmon 1984; Craver 2007: Introduction).

The CM model (Salmon 1984, ch9) tries to capture the elements of a causal explanation over and above statistical relevance. The notion of cause is conceived here as the interaction between causal processes (Dowe 2000), rather than as a statistical notion. A causal process, as opposed to a pseudo process, is a physical process characterized as the ability of transmitting a mark (a

[6] This distinction will be adressed in section 1.2.2

modification on the structure of the process) in a spatio temporal continuous way. For instance, the movement of a baseball through space is a causal process, for from its initial spatio temporal position to its final position, a scuff in the surface of the baseball is transmitted; another paradigmatic example of a causal process is a car in motion, where the mark, in form of a dent in a fender, will persist from one spatiotemporal location to another (Woodward 2017, pp. 36-36). This ability of transmitting a mark is what distinguishes causal processes from pseudo-processes. For instance, the shadow of a moving physical object would be a pseudo-process. If we try to mark the shadow by modifying its shape, the modifications would not be maintained unless we intervene to keep it[7].

Different causal processes can interact with each other. Salmon entitles this *causal interactions*. In a causal interaction, two or more causal processes interact resulting in a change in the structure of the processes involved, having new features they would not have had without the interaction. Following with the car example, a collision between two cars is a paradigmatic causal interaction, for after the accident the cars got dented, changing their structure.

Thus, according to the CM model, explaining an event *E* is to track the causal processes and causal interactions involved in the occurrence of *E*. Explaining an event is then to show how it fits into a causal nexus[8] (Salmon 1984, ch. 9). The account is well illustrated by a pool game, in which different balls move and collide with each other. For Salmon, describing the processes and interactions of the balls explains their subsequent motions. However, this might pose a problem for those explanations in which there is no "action by contact" (Woodward 2017, p. 36)

Critics of the CM model point out that even in cases of "action by contact" there is not a clear distinction between causal processes and causal interactions that explain the event, that is, the *relevant ones*, and other causal processes that do not have explanatory force even though they appear as being part of the causal story of the event (Woodward 2017). For instance, a mosquito hitting the fender of a car few minutes before it had an accident constitutes a causal interaction although it is explanatory irrelevant for the accident (event). Thus, the connection between the *explanandum* and the relevant causal nexus is unclear (Kitcher 1989; Woodward 1989). Salmon, in the attempt to overcome some of these criticisms, changed the "mark transmission" concept of a causal process by the transmission of a non-zero amount of a conserved quantity (Salmon 1994), being such conserved quantities the ones found in physics, say, linear momentum, angular momentum, charge, etc. Nevertheless, Woodward (2017, p. 37) claims that this modification still does not solve the problem of capturing the elements of the causal nexus that are explanatory relevant. For instance, a billiard ball in motion can transmit different conserved quantities, like linear momentum, angular momentum, charge… and it will do so in a collision with another balls. However, Woodward (2017, p. 37) wonders what allows us to pick linear momentum rather than other conserved quantity as the explanatory relevant for the subsequent motion of the ball. The CM model does not provide any answer. Furthermore, cases in which the property explanatorily

[7] As Woodward points out, the ability of transmitting a mark is not a necessary condition for a causal process. The notion is understood counterfactually, as "if a causal process were marked, the mark would persist through time".

[8] Salmon distinguishes between tracking the causal processes and causal interactions leading up to an event E (etiological aspect of the explanation) and describing the causal processes and causal interactions that make up E (constitutive aspects of the explanation). (Woodward 2017, p. 36)

relevant seems not to involve the exchange of a conserved quantity, like "having ingested birth control pills", or "being pregnant", are even harder to consider under the CM model (Woodward 2017, p. 44).

Furthermore, it looks like the CM model might have problems to account for the explanations of the behaviour of complex "higher level systems", in which we encounter "explanations that do not explicitly cite spatio-temporally continuous causal processes involving transfer of energy and momentum" (Woodward 2017, p. 39). Biology and psychology are disciplines that exhibits a great amount of this type of explanations.

There have been other attempts to construe philosophical models of scientific explanation with universalist aspirations. For instance, Philip Kitcher provides his own account of explanation as unification (Kitcher 1989). Kitcher's account could be seen as another attempt to answer question 1 (seeking for necessary and sufficient conditions of a scientific explanation). Briefly speaking, Kitcher believes that one of the most important features of science is its ability to unify phenomena that looked disconnected at a first sight. Theories developed by Newton, Darwin and Maxwell are some examples of unquestionable unifications in science. According to Kitcher, explaining a phenomenon consists in showing how it can be derived from an argument pattern that belongs to the "explanatory store". By explanatory store, Kitcher means a group of arguments that maximally unify (i.e. that maximizes the combination of simplicity and strength) the set of beliefs that are accepted in a specific time by science. Given that there are many ways of systematizing a body of accepted beliefs, Kitcher believes that a systematization will be more unifying than another if it derives a higher amount of conclusion, if its patterns are more stringent and if it uses a smaller number of patterns. An inference will thus be explanatory if it belongs to the best (more unificatory) systematization.

Kitcher's model has never had much acceptance within the philosophical community. It raises several concerns. One of them is the issue of fitting causal relationships with explanatory unification. Another is the notion of unificatory power, which is very vague as Kitcher himself acknowledge. For instance, there are situations in which it would be difficult to tell which systematizations have higher unificatory power: one that derives higher number of conclusions, but uses higher number of patterns, or another with lower derivations but a lower number of patterns? A different worry concerns temporal symmetry, like cases challenging Hempel's DN model as well. Although Kitcher's model may fix some of the counterexamples to Hempel's DN model and IS model (the problem of explanatory relevance, and the problem of asymmetry, see Díez & Moulines 2008), it cannot fix them all. For instance, consider Newtonian mechanics applied to the Solar system. We could explain the further state of motion of planets at a time t considering information about positions at time t_0, masses, velocities, the forces affecting on them at t_0 and the laws of mechanics in a predictive way. However, we could consider *retrodictive* derivations (Woodward 2017) in which the present motion of the planets are derived from information about the future velocities and positions at t, forces operating at t and laws of mechanics. It looks like the pattern used in the retrodictive derivations is as simple and as unified as the pattern used in predictive derivations. Nevertheless, only one would be considered explanatory, while the other would not (Woodward 2017)[9].

[9] Cf. Diez 2014 for a possible reply to this point.

After considering a very brief detour towards three of the most influential philosophical accounts of scientific explanation, one thing is certain: there is no consensus regarding the analysis of a scientific explanation. The identification of necessary and sufficient conditions for an explanation to be genuinely scientific seems to be a really difficult goal to achieve.

Pierre-Alain Braillard and Christophe Malaterre (2015; introduction) suggest that there are two possible strategies to follow at this point. We can pursue the search of an even better philosophical model of scientific explanation, maybe one that develops a better account of causation, or a mix of different notions as the fundamental ones for characterizing scientific explanation. This strategy would still be looking for the aforementioned necessary and sufficient conditions that makes an explanation genuinely scientific, and it would still be guided by replying to the problems 1) to 3). We could call it the monist strategy. A particular type of normativity underlies such a project: the aim of dictating how a genuine scientific explanation *must* be. Accordingly, undergoing the monist strategy would imply the belief that scientific explanations are of a single sort. Although some attempts to this directions have been made (e.g. Díez 2014; Strevens 2004), a closer look at the literature on scientific explanation suggests that the different explanatory strategies used nowadays in science are at odds with the monist aspirations (Mitchell 2003; Kellert et al. 2006; Brigandt 2010, 2013a; Braillard & Malaterre 2015).

The other strategy is to give up on problems 1) to 3) and to the monist attempts and accept that the science provides examples of different kinds of explanations, thus accepting an explanatory pluralism in scientific explanation (Mitchell 2002, 2003; Brigandt 2010, 2013a, 2013b; Braillard & Malaterre 2015). This second strategy seems to be more promising. It is a less ambitious project given that it does not have normative motivations or at least it does not have the monistic ambitions of the traditional accounts of explanation. This second strategy puts scientific practice in the foreground, and analyses how scientists explain in different branches, without imposing much constraints in how explanation must work. Moreover, this later strategy fits better with the current research in philosophy of science, where investigations tend to be less general and much more focused in a specific domain, say, biology, physics, psychology, and in sub-branches within these disciplines, molecular biology, quantum physics, cognitive psychology, etc.

Even if nowadays the pluralist stance seems to be more present and prominent in the philosophy of biology –as e.g. the book edited by Braillard and Malaterre (2015) contends– there are several questions that a pluralist in scientific explanation should address, like what is the relation between the different models of explanation (section 1.4 will address this and other questions).

Moreover, and in addition to the monism-pluralism debate, Braillard and Malaterre (2015, introduction) contend that the current debates on scientific explanation in biology could be seen as attempts to reply to the following four questions:

I. Are there natural laws in biology?
II. Does causation play a specific explanatory role in biology?
III. Are there other types of explanation needed?

IV. Is the New Mechanism strategy, which seems to integrate some kind of law-like generalizations with causation, the unique model of explanation in biology?[10]

The three papers that conform this thesis dissertation should be read as a contribution to the pluralist strategy. The papers do not aim at defending a monist account of scientific explanation. Rather, what they do is to put some limits to the universalist aspirations of the widely accepted and highly successful New Mechanism account of scientific explanation in biology. They do so by suggesting two complementary explanatory accounts that are needed to analyze the explanatory power of some biological models, suggesting thus an explanatory pluralism in biology. Furthermore, the papers contribute to answer question II and IV mentioned above, although by addressing these questions, they display arguments that tangentially reply to questions I and III. They do so by examining three different case studies in contemporary biology to show that even if they could be though *a priori* to fit with the New Mechanism model of explanation, a closer analysis shows that this is not the case. Given that each of the three papers departs from the New Mechanism model of scientific explanation I will devote next section to characterize it in detail.

1.2. New Mechanism account of scientific explanation

The most widely accepted philosophical account of scientific explanation in biology nowadays is the one provided by New Mechanism philosophy, namely mechanistic explanation (Bechtel & Richardson 1993; Bechtel & Abrahamsen 2005; Machamer, Darden & Craver 2000, MDC henceforth; Craver 2007; Glennan 2002, 2017; Glennan & Illari 2018). A mechanistic explanation can be seen as a type of CM explanation (Salmon 1984), for it claims that mechanisms are explanatory in virtue of their capacity to cause the phenomena that are being explained (Craver & Tabery 2015). The New Mechanism account of scientific explanation has had a major impact in contemporary philosophy of biology than Salmon's account, since it focuses on the practice and models of biologists, rather than on physics, which is the domain that better fits Salmon's CM model of scientific explanation (Glennan 2017, ch. 1). In order to introduce mechanistic explanation, I must outline the main features of this New Mechanism philosophy. Although New Mechanism philosophers are cautious in claiming that explanations in biology are all causal-mechanistic, in the literature they usually confront mechanistic explanation with Hempel's model, as if mechanistic explanation were to take the dominant position that years ago Hempel's model took, at least in biology. Others even go further and claim that we could extend the talk of mechanisms along other sciences (Illari & Williamson 2012; Glennan 2017)

The New Mechanism philosophy has had and is having a major impact in the philosophy of science and, particularly, in the philosophy of biology for the last 20 years. The New Mechanism philosophy provides a framework to think in a different way about some of the most important topics in the philosophy of science, like laws, causation, function and explanation (MDC 2000). Stuart Glennan, at the very beginning of his recent book about New Mechanism philosophy

[10] Braillard and Malaterre do not exactly ask in question IV whether New Mechanism account of explanation is the only philosophical model of scientific explanation in biology. They say whether it "fulfils all expectations". Given the big success New Mechanism philosophy is having in contemporary philosophy of science, and the aspirations of extending mechanistic talk along other science beyond biology (e.g. Stuart Glennan), the question is though relevant.

(Glennan 2017) states that according to New Mechanists "most or all the phenomena found in nature depend on mechanisms", and that science's chief business is "the construction of models that describe, predict and explain these mechanism-dependent phenomena" (Glennan 2017, p.1). So, New Mechanism is ultimately a view about nature and science, and not only a thesis about explanation. This is the main reason why I need to clarify the commitments of the New Mechanism philosophy before discussing their particular account of scientific explanation.

Holly Andersen (2014) and Arnon Levy (2013) argue that we should distinguish different theses and claims within the New Mechanism philosophy. Andersen distinguishes three main claims at the roots of New Mechanism. The *ontological claim* states that the phenomena studied by many sciences (particularly by those involving the study of higher-level phenomena like biology) have a mechanistic-type and hierarchical structure, each level involving entities organized in specific ways and connected *via* causal interactions. The *methodological claim* supposes that the methodology of many sciences reflects the mechanistic ontology of the phenomena under study, by "isolating consistent patterns or regularities in the world and decomposing them into constituent entities, causal connections and spatio-temporal organization" (Andersen 2014, p. 275). The third claim, *explanatory claim*, tells that explanations in these fields reflect the mechanistic ontology, showing how particular phenomena are the end product or "constituted by operation of such mechanisms" (Andersen 2014, p. 275). In the same vein, Levy (2013) distinguishes between three main theses. *Causal mechanism*, a particular account of causation claiming that causal relations exist in virtue of underlying mechanisms. *Explanatory mechanism,* a thesis regarding the nature of explanation, in which to explain a phenomenon is to cite causal-mechanistic information. *Strategic Mechanism,* a thesis that contends that some phenomena are better handled mechanistically, highlighting the advantages of modelling in such a way.

Moreover, according to some philosophers of biology (Craver & Darden 2005; Nicholson 2012; Glennan 2017, ch. 1) the notion of *mechanism*, the essential concept in the New Mechanism philosophy, might be confusing. That is why it is important to clarify this notion in order to avoid further confusions. Nicholson (2012) for example, claims it is important to distinguish three different meanings of the term "mechanism" in biology, and argues that without that distinction several works in the New Mechanism tradition suffer from a conflation of meanings that leads to confusions. Nicholson distinguishes between a) *mechanicism*[11], the philosophical thesis that conceive living organisms as machines, so a philosophical thesis about the nature of life (Galileo, Descartes…); b) *machine mechanism,* the internal workings of a machine-like structure (Militello & Moreno 2018); and c) *causal mechanism,* a particular type of explanation.

Having these distinctions in mind, I now turn to examine the concept of mechanism as used by the New Mechanism Philosophy. As we will see, this concept is essential to properly characterize the mechanistic account of explanation.

[11] Glennan contends that *mechanicism* is at the roots of the New Mechanism philosophy but highlights important differences. For instance, the *mechanicist* view, opposed to vitalism, is committed to atomism (metaphysically and methodologically), while New Mechanism is not. *Mechaniscists* of the seventeenth-century were committed to the idea that all phenomena could be explained "in terms of action by contact of variously shaped microscopic corpuscles" while New mechanists do believe that explanations can be given by the appeal to objects of different kind and size (Glennan 2017, p. 6).

1.2.1.Mechanisms

To properly characterize mechanistic explanation -and given that according to the New Mechanism philosophy to explain a phenomenon is to identify the mechanism responsible for its appearance- I need to provide a minimal definition of what a mechanism is. The following three definitions are some of the most cited and are generally assumed to capture the essential features of a mechanism (Illari & Williamson 2012; Tabery & Craver 2015):

(i) Machamer, Darden and Craver: "Mechanisms are entities and activities organized such that they are productive of regular changes from start or set-up to finish or termination conditions." (MDC 2000, p.3)

(ii) Glennan: "A mechanism for a behaviour is a complex system that produces that behaviour by the interaction of a number of parts, where the interactions between parts can be characterized by direct, invariant, change-relating generalizations." (Glennan 2002, p.S344.)

(iii) Bechtel and Abrahamsen: "A mechanism is a structure performing a function in virtue of its component parts, component operations, and their organization. The orchestrated functioning of the mechanism is responsible for one or more phenomena." (Bechtel and Abrahamsen 2005, p.423)

The three definitions point out towards four essential and shared features of a mechanism. Phyllis K. Illari and Jon Williamson (2012) refer to these features as the *phenomena*, the *entities* and *activities* (or *parts* and *operations*), and their *organization*. Carl Craver and James Tabery (2015), on the other hand, refer to them as the *phenomenon*, the *parts*, the *causing* and the *organization.* In what follows I briefly characterize the basic features of a mechanisms: the *explanandum phenomenon*, *entities and activities*, their *organization*, and the *causal nature* of the mechanism.

Phenomenon

Everyone in the New Mechanism philosophy believes that a mechanism always has the capacity of producing a phenomenon. As Justin Garson expresses, "the phenomenon that a mechanism serves is not somehow incidental to that mechanism, but constitutive of it: mechanisms are identified and individuated by the phenomenon they produce" (Garson 2018, p. 104).

The phenomena under study, i.e. the biological phenomena demanding an explanation, can be of a different sort. In general, New Mechanists talk about a mechanism producing, underlying, or maintaining a phenomenon. For instance, the phenomenon under study can be seen as the end state or the end product of the mechanism, i.e. its result, being this a behaviour, a capacity or an object. The mechanism of protein synthesis has as its end product an object, namely a specific protein (Darden 2008). In other cases, for instance in the case of physiological mechanisms like the neuron action potential, it is better to interpret the phenomenon as a behaviour, e.g., the behaviour of the mechanism to generate a change in the electrical potential difference across a neuron's membrane (Craver 2006). In yet other cases it is better to see the mechanism as maintaining the phenomenon, like homeostatic mechanisms that hold body temperature between certain boundary conditions.

Another dimension concerns *regularity*. There seems to be disagreement on whether there must be a regular relation between a mechanism and a phenomenon. There are New Mechanists criticizing MDC's definition of mechanism precisely due to its requirement of regularity. It has been claimed that this feature excludes cases that occur only once (like the transition from prokaryotic to eukaryotic cell), and that something must work more than once to be considered a mechanism. However, according to Craver, the regularity claim could be interpreted as a counterfactual, saying something like "were all the conditions the same, then the mechanism would likely produce the same phenomenon" - "likely", because they consider the possibility of stochastic mechanisms - (Craver & Tabery 2015). In contrast, Krickle (2014), coining the notion of *reverse regularity*, contends that there must be a regularity of the form "the mechanism was acting when the phenomenon occurred" (See Glennan & Illari, 2018, ch 7, for a classification of mechanisms given the regularity they instantiate).

Entities and activities

As William Bechtel and Robert C. Richardson (1993) taught us, mechanisms' decomposition is an important goal of contemporary biological research. Broadly speaking, entities and activities are the two main constituents of a mechanism. On the one hand, mechanisms are composed by entities (also called parts or component parts). On the other hand, entities act in a certain way, performing activities and interacting among them to produce the phenomenon, and thus eventually producing changes in a system (MDC 2000). For Glennan (2018, ch. 2), whose notion of mechanism is minimal to embrace as many domains of science as possible, entities are objects, such as organisms, protein chains, neurotransmitters, muscle fibers, genes, and even congressional committees and planets. Glennan contends that entities have stable properties and boundaries. Moreover, Craver provides four criteria that the parts or entities of a mechanism satisfy: they realize a stable cluster of properties, they are robust, they can be used for interventions and they are physiologically plausible in a given pragmatic context (Craver 2007, ch. 4).

Moreover, the activities require that the entities have certain properties allowing them to engage in actions. For instance, a DNA base and its complementary hydrogen base bond together because they have weak charges and a specific geometrical structure (MDC 2000). Activities include patterns as diverse as bounding (of atoms), transcription (of DNA), folding (of a polypeptide chain into a functional protein), contracting (muscle fibers), etc.

There is currently a debate about the metaphysical status of the entities and the activities. MDC are dualists, claiming that mechanisms are made of two different kinds of components: entities and activities. Glennan, on the contrary, argues the opposite, claiming that the unique components of mechanisms that have a metaphysical reality are entities, while interactions and activities are but "occasions on which a property change on one entity brings about a change in properties of another entity" (Illari & Glennan 2018, ch. 9, p. 116).

Organization

In a mechanism, entities and activities are organized in such a way that they are productive of the phenomenon. For instance, in an electrical circuit, you can have the same resistors generating different resistance whether they are wired in parallel or in series. Craver (2007, ch. 4) contrasts organization with mere aggregation. In an aggregate system, the whole is the sum of its parts.

Suppose that we explain the property or behaviour Ø of a system S by the properties and activities of its parts. In an aggregate system, the Ø will remain invariant even if we rearrange the parts and even if we add or subtract parts. Besides, in an aggregate system there are no interactions between the parts that are relevant for Ø (Craver 2007, ch. 4). Mechanistic components are however arranged in a way that is opposed to the way aggregate systems are. This is because if you undergo the types of changes described above in a mechanistic system, the phenomenon Ø will definitively change. In a mechanism, entities interact in specific ways with one another, and for the phenomenon produced by the mechanism, it matters which entity interacts with which others. Thus, rearrangements, subtractions or additions of entities will modify the way the entities interact, affecting the final outcome Ø. According to Craver (2007, ch. 4) the organization mechanisms instantiate is called "active organization", and he contends that no mechanism is complete until it captures such kind of organization. Active organization is sustained by the spatial and temporal organization of the components of the mechanism. By spatial organization, Craver means the location, size, shape, position and orientation of the component parts, while by temporal organization he refers to order, rate and duration of the activities (Craver 2007: ch. 4; Craver & Tabery 2015). Craver claims that these types of organization are differently weighted depending on the mechanism under study. In biochemical cascades, for instance, the location of the activities is less important, while the structure given by the entities and the temporal arrangement of the activities is much more important.

Causation

The last essential feature of a mechanism is its causal nature. A mechanism can be seen as a sequence of different *entities* connected by different *activities* organized in a *causal way* that end up with the realization of a phenomenon (object, behaviour or capacity). However, New-mechanists are far from a consensus about how to interpret the nature of these causal relations. Craver and Tabery (2015) believe there are four ways New Mechanists talk about causation, all having in common the rejection of the Humean regularitivist account embraced by logical empiricists. They classify these different ways as *conserved quantity accounts*, *mechanistic accounts*, *counterfactual accounts* and *activity-based accounts*. I will briefly present the mechanistic and the counterfactual accounts for they seem to be the most used by New Mechanists[12]. The mechanistic account, mainly embraced by Glennan (1996, 2017, ch. 6), contends that causation is derived from the notion of mechanism itself. According to this picture of causation, causal claims are about the existence of an underlying mechanism that connects cause and effect

This account has been charged of circularity. In trying to provide an exhaust definition of a mechanism, one has to speak of causation, and to define causation, one has to speak about the mechanism, so that this opens a question about fundamentality (Craver & Tabery 2015). Glennan tries to solve this problem claiming that at least in higher-level causes, rather than in fundamental causes, a mechanism always explains the connection between cause and effect (Glennan 2018, ch. 6). Thus, the viability of this account depends on how you treat the circularity problem (Craver & Tabery 2015).

[12] Nevertheless, I believe that the most used and successful account of causation, mainly embraced by Craver and by those who work with mechanisms as means to reach a definition of scientific explanation, is the counterfactual account based on Woodward's manipulationist theory (see Craver and Tabery 2015 for a detailed picture of these accounts)

On the contrary, the view of causation I believe fits better with mechanistic explanation and the New Mechanism philosophy is the manipulationist account of causation developed by James Woodward (2003). This view is attractive for it also provides an analysis of explanatory relevance, which is fundamental for any account of causal explanation in science. The basics of Woodward's account is that causal relations are relations potentially exploitable for purposes of manipulation and control (Woodward 2016). So, manipulation on a cause will result in the manipulation of its effect: in other words, "if C is genuinely a cause of E, I can manipulate C in the right way, this should be a way of manipulating or changing E" (Woodward 2016, p.2). The manipulation of C is always performed by an intervention. This account is considered a difference-making account, for it tells that identifying a cause of some effect is to find an event or property that is making a difference regarding another event. If you can modify that event via an intervention, you will definitely get a modification of its effect. Manipulation via interventions though is not always possible (think about cosmology or past events), so this account has been charged for providing a picture of causation too anthropocentric. It has been charged as well of circularity, for you define interventions in terms of causation and causation in terms of interventions (Woodward 2016). In general, interventionists reply to the first worry by claiming that we could use manipulation as an heuristic tool, so as to build counterfactuals of the sort, "if you were to manipulate X, it would follow a change in Y", where the story told (an ideal intervention) is what provides the means for identifying the causal relationship. The accusation of circularity is solved in different ways, but for the purpose of illustrating the causal nature of mechanisms we do not need to go deeper on this problem (see Woodward 2002, 2003 as a way out to circularity). As advanced, this account provides an analysis of explanatory relevance, for ideal manipulations serve to test the relevance of an entity or part with its performed activity for the production of the phenomenon. Besides, it is ontologically minimal for when you claim that C causes E you need to show how an ideal intervention in C would produce a change in E, rather than telling that C and E are physically connected.

Having clarified the notion of mechanism by the New Mechanism philosophy, I now turn to characterize the notion of mechanistic explanation.

1.2.2. Mechanistic explanation.

Mechanistic explanation is a type of causal explanation. We can distinguish two types of causal explanation, etiological and constitutive (Salmon 1984; Craver 2007; Halina 2018). An etiological causal explanation tells about the causal story of the explanandum phenomenon, it cites (some of) its antecedent causes in a general way, like when someone generally says a virus explains a disease. A constitutive causal explanation, on the contrary, explains a phenomenon by describing the mechanisms underlying it (Craver 2007; Halina 2018; Glennan 2017). Craver contends that Salmon and other philosophers of science that have analyzed causal explanation in science have always been focused on etiological causal explanations. Mechanistic explanation thus should be called constitutive mechanistic explanation, for it focuses on the specific causal mechanistic details of a phenomenon.

A constitutive mechanistic explanation of a phenomenon consists on the description of the mechanism, with its entities, activities and organization that is responsible for the production of (or maintaining) the phenomenon. The explanation must show how the entities of the mechanism organized in a specific way causally produce the phenomenon through the activities performed. Thus, a mechanistic explanation shows how the production relation between phenomenon and the

constituents of the mechanism is causal and as stated above, these causal relations are understood in a manipulationist way, following Woodward (2003).

When scientists provide mechanistic explanation what they do is to build explanatory models, in this case, we will talk about models of mechanisms. Mechanisms, New Mechanists claim, are real things, composed of real parts and engaged in real activities. The way scientists formalize knowledge about such mechanisms is by building models of mechanisms, a tool they will use to achieve the pragmatic goals of science: explaining, intervening and representing (Mitchell 2000). Tarik Issad and Christophe Malaterre (2015) contend that a mechanistic explanation has two main parts. On the one hand we have a model of a mechanism, and on the other we have a causal story telling about how the entities are causally connected among them, instantiating certain activities, and with certain organization, which eventually lead to the production of the phenomenon. Issad and Malaterre (2015) built five "normative elements" based on Craver's considerations (Craver 2007, p. 161), that a model of mechanism must fulfil in order to contribute to a mechanistic explanation:

(i) the model of mechanism must account for all aspects of the *explanandum* phenomenon,
(ii) the model of mechanism must be based on components and activities that are real[13]
(iii) these activities must be causal [in a manipulationist sense]
(iv) the entities and activities are spatially and temporarily organized
(v) the entities and activities are all relevant with respect to the *explanandum* phenomenon

As mentioned, Issad and Malaterre believe that a model of mechanism *per se* fail to be an explanation without a further causal story. They believe, and I agree, that a list of entities and activities even with their active and spatio-temporal organization is rarely explanatory. A mechanistic explanation is complete when beside the model of mechanism, a causal story is provided, a causal story that tells how the entities and activities causally interact so as to produce the *explanandum* phenomenon. Issad and Malaterre speak about a rehearsing of the unfolding of the activities of the different entities, like if we "were to run through all of the different causal links in the right sequential order" (Issad & Malaterre 2015, p. 269). They illustrate this rehearsing with a canonical example from neuroscience due to Craver (2007), explaining the movement of Na+ ions through the neuronal membrane during the action potential:

> At Vrest [the membrane potential at rest], a positive extracellular potential holds the α -helix [a subunit of a trans-membrane molecular complex known as the Na+ channel] in place. Weakening that potential, which happens when the cell is depolarized, allows the helix to rotate out toward the extracellular side (carrying a "gating charge" as positively charged amino acids move outward). This rotation, which occurs in each of the Na+ channel's subunits, destabilizes the balance of forces holding the channel in its closed state and bends the pore-lining S6 region in such a way as to open a channel through the membrane. Another consequence of these

[13] The distinction advanced in MDC (2000) between schemas and sketches might illustrate the accuracy or truth of models of mechanisms. On the one hand schemas are abstract descriptions of mechanisms that can be filled in later if needed to specify type or token mechanisms. A sketch is an incomplete description of a mechanism that captures some of the entities and activities but it has black boxes that cannot be filled in yet. How possible and how actually-enough schemas, on the other hand captures the hypothetical character of mechanisms. A how-possible schema describe how entities and activities could be organized, though this model could be false. A how-actually-enough model of mechanism is a true how-possible model of mechanism (Craver & Tabery 2015).

> conformation changes is that the pore through the channel is lined with hairpin turn structures, the charge distribution along which accounts for the channel's selectivity to Na+. (Craver 2007, p. 119).

This passage provides the causal story, essential for a mechanistic explanation. It provides the workings inside the mechanism that produces the phenomenon, the flux of Na+ ions generating an action potential. The passage talks about entities ($\alpha - helix$, hairpin turn structure) and activities (holding in place, rotating, bending) arranged in a specific spatio-temporal way. According to Issad and Malaterre, this is the typical causal story needed to complement the model of mechanism and build a mechanistic explanation (Issad & Malaterre 2015, p. 270)[14].

To sum up, following Issad and Malaterre (2015, p. 270) "a mechanistic explanation (ME) explains a phenomenon (P) produced by a system (S) in virtue of fulfilling the following two necessary and sufficient conditions:

> (MM) Displaying a model of mechanism that represent a real mechanism in S with its entities, activities and spatio-temporal organization
>
> (CS) Rehearsing a causal story that enumerates the cause-effect relationship taking place in the mechanism up to the production of the phenomenon P".

A final comment needs to be made regarding causal relevance. All accounts of causal explanation need to address the problem of finding a way to distinguish relevant from irrelevant causal information for the explanation of a phenomenon. In mechanistic explanation, given that very often models of mechanisms describes complex systems, there must be a way of identifying relevant and irrelevant factors of a mechanism responsible for the production of the phenomenon. The most successful way to deal with this problem is to use Woodward's manipulability account (mentioned above). Woodward contends that it is by experimental or ideal manipulations that one discovers about the constitutive relevance of a factor for the production of a phenomenon (see Craver 2007, ch. 4, for a detailed discussion of it).

1.3. Limits of mechanistic explanation

It's hard to underestimate the success of mechanistic explanation in capturing the nature of scientific explanation in biology. The impressive amount of literature generated by New Mechanism philosophy and in particular about mechanistic explanation is also hard to ignore (see Craver & Tabery 2015; Glennan & Illari 2018). New Mechanism philosophy has contributed to different debates within the philosophy of science, such as those about the metaphysics of causation, the relation between laws, counterfactuals and mechanisms, emergence, reduction,

[14] Other paradigmatic examples of mechanisms and mechanistic explanations are the mechanism of protein synthesis (Darden & Craver 2002), the mechanism of the Krebs cycle, a step in the metabolism of sugar (MDC 2000, Craver & Darden 2013, ch.2), the mechanism of cell reproduction, cell respiration, (Bechtel & Richardson 1993), fermentation as a metabolic process (Bechtel & Abrahamsen 2005), the mechanism of heredity (Craver & Darden 2013, ch. 10), to mention just a few.

natural kinds and others (see Glennan and Illari, 2018, as a compendium of papers about different applications of New Mechanism philosophy).

The mechanistic framework has proved very successful at the task of illuminating the nature of explanations in experimental biology, such as molecular biology, physiology, cell biology, genetics and neuroscience, among others (see Craver 2007; Brigandt 2013b). Those branches of biology are mechanistic in nature, and the process of describing mechanisms and decomposing them in sub-parts is common currency in the scientific practice (Becthel & Richardson 1993).

However, several authors have argued for a restriction of the scope of mechanistic explanations in biology (Issad & Malaterre 2015; Mekios 2015; Théry 2015; Brigandt 2015; Brigandt et al. 2018; Halina 2018). As mentioned in section 1.2, New Mechanists do not claim all scientific explanations are mechanistic. Nevertheless, a closer look at the literature reveals that there's strong inclination towards the view that specifically in biology, scientific explanations are mechanistic. For instance, Robert A. Skipper and Roberta L. Millstein (2005) provide an interpretation of evolutionary theory in causal mechanistic terms even though other theorists contend that evolutionary explanations are statistical and non-causal, as opposed to mechanistic, in kind (Brigandt 2013a; Walsh 2015; Walsh et al. 2002, 2017).
In this section I will focus on two lines of criticism that put some limits to the wide scope of mechanistic explanation in biology.

First, it can be argued that the role that mathematics plays in several biological explanations weakens mechanistic explanation in some quantitative models. New Mechanists in general consider mathematics to play only a representational or descriptive role in biological models. In other words, mathematical models in biology, for instance dynamical models, are seen as tools for describing the workings of a complex system (Kaplan & Craver 2011; Kaplan & Bechtel 2011). According to New Mechanists the only way mathematics can play an explanatory role in a biological explanation is by revealing something about the causal structure of a mechanism. However, several authors who do not belong to the New Mechanism movement (Chemero & Silbernstein 2008; Stepp et al 2011; Huneman 2010, 2018 a,b,c; Jones 2014; Issad & Malaterre 2015) claim that the explanatory force of some mathematical models in biology does not lie on the description of constitutive causal- mechanistic information; rather, such models show how the biological system under study can be seen as having a particular mathematical structure, with certain mathematical properties that account for the phenomenon to explain (Huneman 2018a). Thus, the challenge here is to show, contrary to what New Mechanists claim, how certain mathematical models in biology are explanatory in virtue of capturing certain structural features of a system and not for identifying posits of the causal structure of a mechanism (Halina 2018).

Secondly, New Mechanists believe the talk of laws of nature is useless in biology nowadays (MDC 2000; Craver & Kaiser 2013). Instead they claim the regularities found in nature are captured by the notion of mechanism itself (Bechtel & Abrahamsen 2005; Craver 2007). However, some philosophers (Carrier 1995; Brandon 1997; Leuridant 2010; Mitchell 2003; Dorato 2012; Díez & Lorenzano 2013, 2015, Alleva et. al 2017), replying to Beatty's Evolutionary contingency thesis (1995), still believe that we can talk about laws in biology in non-mechanistic terms, and give them an important role in some explanations in biology. Following this strand of thought, some philosophers believe that Hempel's thesis to the effect that explaining a phenomenon is to make it expectable on the basis of lawful conditions, is worth keeping in scientific explanation. For instance, José A. Díez (2014) elaborates a Neo-Hempelian

covering law model of scientific explanation (ASE, standing for Ampliative Specialized Embedding). This account, which has its roots in the structuralist tradition in the philosophy of science (Balzer et. al 1987) shows that there are other ways an explanation in biology gains its explanatory force. Instead of gaining it by appealing to a model of mechanism and a causal story, the neo-Hempelian account appeals to laws of nature and to an ampliative component as a theoretical novelty to explain (more details in section 1.3.2).

1.3.1. Mathematical explanation in biology

The explanatory role that mathematics and mathematical models plays in biological explanations has been a topic of discussion over the last years. The New Mechanism position that mathematical modelling is merely a representational tool for describing a complex system, together with the idea that mathematics in a biological model are only explanatory if they capture part of the causal structure of the mechanism (Kaplan & Bechtel 2011; Kaplan & Craver 2011) is under discussion. The rise of systems biology[15], to be regarded as a new interdisciplinary field combining elements from molecular biology and physiology with quantitative modelling in mathematics or computer science, puts mathematical models under the spotlight (Green 2017, 2019; Brigandt et. al 2018; Kaplan 2018).

A dynamical system, a notion which features, for instance, in systems biology and in neuroscience among others, can be seen as a set of variables interacting over time. When those variables change over time, they generate different patterns (Kaplan 2018). Modelling these systems is usually done with differential equations "which specifies how a system changes at any given time point as a function of its state at that time" (Kaplan 2018, p. 268). Thus, given the recent scientific interest in dynamical systems modelled mathematically, philosophers of science wonder about the kinds of explanations those models provide. We can identify two different positions at this stage.

On the one hand, there have been different attempts to extend the notion of mechanistic explanation so as to integrate mathematical modelling in a compelling way (Kaplan & Bechtel 2011; Kaplan & Craver 2011; Brigandt 2013b, 2015; Kaplan 2018). In particular, William Bechtel and Adele Abrahamsen have developed the notion of *dynamic mechanistic explanation* as an attempt to integrate the explanations provided by mathematical models within a mechanistic framework (Bechtel & Abrahamsen 2010, 2011, 2012, 2013). According to such an account, a dynamic mechanistic explanation focuses as well on concrete entities and interactions but extends mechanistic talk with mathematical models to account for "the dynamical operations of the system across time" (Bechtel & Abrahamsen 2010). Drawing on neuroscience, Bechtel and Abrahamsen explain how mathematical models, together with models of mechanisms, explain the dynamics of large systems of non-linear organization (like circadian rhythms) in a causal mechanistic way (Bechtel & Abrahamsen 2010). In a *dynamic mechanistic explanation*, even if the mathematical modelling is essential, the explanation still gains its explanatory force in virtue of the identification of causal mechanical information. Others, like David M. Kaplan and Carl Craver (2011) or Kaplan and Bechtel (2011), challenge detractors of mechanistic explanation to provide an alternative model of explanation that presents in detail how mathematical models explain without capturing the causal structure of a mechanism.

---

[15] Systems biology is a recent integrative and interdisciplinary discipline (Green 2017). Its main goal is to explain the workings of complex biological systems using experimental and conceptual tools. According to Brigandt, systems biology focus on "system wide-behaviour rather than the properties of few isolated components or causal pathways of a system" (Brigandt 2013b, p. 478).

On the other hand, a different strand of research (Batterman 2002; Chemero & Silbernstein 2008; Stepp et al 2011; Batterman & Rice 2014; Ross 2015) believes that mathematical models should be seen as an alternative explanatory stance to the New Mechanism picture of explanation in biology. These philosophers question the possibility of integrating mathematical modelling, like ordinary differential equations or network modelling, with mechanistic explanation. Focusing mainly in neuroscience those philosophers challenge the thesis, defended by Kaplan and Craver (2011), that the explanatory role of dynamical models increases as they include more relevant causal mechanistic details. For instance, Anthony Chemero and Michael Silbernstein (2008) and Nigel Stepp, Chemero and Michael T. Turvey (2011) rely on the ability of those mathematical models to predict and describe, claiming that those features are enough to call them explanatory, whereas Kaplan and Bechtel (2011) criticize this position. Lauren N. Ross (2015) and Robert W. Batterman and Collin C. Rice (2014) provide examples of mathematical models that do not satisfy Kaplan and Craver (2011) constraints that a mathematical model must fulfil in order to be explanatory, what they call a "model to mechanism-mapping" (Kaplan & Craver 2011). Batterman and Rice (2014) and Ross (2015) appeal to Batterman's *minimal model explanation* (Batterman 2002), as an alternative to mechanistic explanation. Explanations given by those mathematical models, they claim, are important not only for their descriptive accuracy and predictive force, but for the abstraction of the mechanistic details they undertake. In other words, those minimal models are explanatory not because they have certain features in common with real systems but by showing that heterogeneous microscopic systems display the same behavioural patterns at a macroscopic scale (Batterman & Rice 2014).

Ingo Brigandt, who undertakes an intermediate position (Brigandt 2013b; Brigandt et al. 2018) grants that in some occasions dynamic mechanistic explanations might work in some modelling in systems biology by way of integrating mechanisms and mathematical models. This would therefore show that, in order to provide a full explanation, the appeal to mathematical models, over and above the appeal to mechanistic information, is indeed necessary. On these grounds, Brigandt, just like Bechtel and Abrahamsen (2010, 2011) tries to extend the notion of a mechanistic explanation in order to find a way to properly integrate mechanistic and mathematical explanations (2013b). However, Brigandt claims that we should not hastily conclude that the integration between mathematical models and mechanistic explanation is always possible and suggests that systems biology explains in certain occasions using only equations and quantitative models, instead of appealing to mechanistic information. He argues that, due to the recent success of systems biology, New Mechanists should pay more attention to investigating the significance of equations and mathematical analysis qua explanatory tools (Brigandt 2013b).

Following this trend, Phillipe Huneman, focusing mainly in biology, has coined the term structural explanation so as to cover different types of mathematical explanations, (Huneman 2018a). Structural explanations are explanations in which the mathematical structure of the system under study plays a key explanatory role for the *explanandum* phenomenon, and not just a representational role (Huneman 2018a). Huneman (2018a) claims there are different types of structural explanations in science that explain using mathematics and not following the New Mechanism paradigm. Huneman mentions minimal model explanations (Batterman 2002, Batterman & Rice 2014, Ross 2015), statistical explanations (Lange 2013, Walsh 2015), topological explanations (Huneman 2010, 2018b; Woodward 2013; Jones 2014) and equilibrium explanations (Sober 1983; Kuorikoski 2007, Potochnik 2015). Huneman believes that the works just mentioned provide instances of non-mechanistic and non-causal explanations while, at the same time, allowing for the possibility of integrations. As a consequence, on Huneman's view,

some structural explanations might be causal. For the sake of clarity, in this thesis dissertation I will follow Issad & Malaterre (2015) in considering their two necessary and sufficient conditions for an explanation to be mechanistic (MM and CS, mentioned in 1.2). If only one or none of these conditions are met, the explanation is deemed non-mechanistic (and to a certain extent, non-causal).

According to Huneman (2018a), all tokens of structural explanations have some commonalities:

i) They all aim at accounting for some pattern rather than just detecting patterns in the data (they are not mere representations/descriptions).
ii) The *explanandum* of a structural explanation, being a property of a system, is not explained by the causal details that lead to it. These details are not explanatory for the behaviour of the whole system, they are abstracted away. To illustrate this point think about why "stones left falling on the top of the hill end up in the valley" (2018a, p. 670). The trajectory of each stone does not matter for the *explanandum*, but just the fact that all end up in the same place.
iii) All structural explanations reach a level of generality that is not achievable by mechanistic research, this is why the specific nature of a mechanism does not figure in the explanation (Moreno & Suárez, (submitted). Because of that, one can change the nature of a mechanism and the structural explanation in which the mechanism is involved would still be valid (network models, as we will see, are good examples of that).
iv) Finally, all these explanations use formal features formulated in mathematical terms.

For Huneman (2018a), the role that a mathematical property of a system plays which can be either representational or explanatory, highly depends on the *explanandum* phenomenon. For instance, he claims that very often, mechanisms are described by functions like "$y = f(X_1 \dots X_n)$" or by a set of differential equations $\frac{dx_1}{dt}$. Once this is set, Huneman claims:

> Using f to compute the values of y and therefore understand which outcomes are predictable given a specific range of inputs, makes use of the function f—hence of the many functional properties that f possesses (e.g. differentiability, monotonicity etc.)—in a representational way in order to explain why the mechanisms being what they are accounts for some of the system's features. Inversely, when I consider that f is in a function (sub)class Cf—defined by a few characteristics of f—that possesses specific properties, which implies that the system (whatever the mechanisms are) will display some features because they are entailed by such properties, this makes an explanatory use of the mathematical properties of that functional class. (Huneman 2018, p. 686)

Hence, we cannot determine whether the mathematical properties linked to an equation or a function used to model a phenomenon are explanatory or representational without a careful analysis of its role.

To clarify this matter, this introduction will now focus on topological and equilibrium explanations, both considered types of structural explanations for Huneman.

Topological explanations

According to Huneman, "[a topological explanation] is a kind of explanation that abstracts away from causal relations and interactions in a system, in order to pick up some sort of topological properties of that system and draw from those properties mathematical consequences that explain

the features of the system they target" (2010, p. 214). Network models are examples of modelling that provide topological explanations. In a more schematic way, a topological explanation explains a feature, outcome or property X of a system S by the fact that S possesses topological properties T_i. A system possessing topological properties means that you can represent S in a mathematical space S', and that mathematical space have property T_i (Huneman 2010). S' can be a graph or network, and graph and networks have structural properties (Xia 2010; Bachmaier et al. 2013)[16]. For instance, a network might be random or non-random, it might display certain structure, like scale free, or small worlds (Solé & Montoya 2001), and each of these structures have different consequences with respect to some features (like stability, robustness, etc).

The point that is of import here is twofold: first, what provides explanatory force to a topological explanation is the identification of a topological property in a system S' that represents the systems under study; secondly, the processes and mechanisms happening at that system S do not matter for the explanation of the *explanandum*. However, these causal mechanisms can be explanatory of another *explanandum*, though less general and more specific.

An illuminating and paradigmatic example of topological explanations can be found in ecology by looking at the diversity-stability debate. There is a longstanding debate among ecologists on whether diversity of species in a community fosters ecological stability or otherwise it destabilizes it (Nikisianis & Stamous 2016). Since the definition of stability might slightly vary for the purposes of the scientists, the debate gets even harder to tackle. In ecology, it is widespread to use network modelling to study the stability of a given ecological community. Some scientists, (Sole & Montoya 2001; Strogatz 2001; Huneman 2010) consider three main parameters to build such networks: the number of species and the average number of links per specie, the number of connections realized versus the possible connections, and the distribution of connections between species. According to Huneman, the network treatment of ecological communities provides topological explanations of the stability and robustness of those communities.
The first step in such kinds of models is to build the network S' for the ecological community (system S); the second step is to study (via simulations) the features of the network to derive mathematical properties of it. In these networks, every node represents a species and interactions between these species are represented by the links between nodes. The types of causal interactions/relations between two species can be of different sort: species A can prey on species B and the other way around; A can compete, be mutualistic or commensal with B, or A can be parasitic on B, etc, (Huneman 2010). However, the nature of those interactions in network modelling is irrelevant for the *explanandum* behaviour (stability of the community). What matters is the final shape or structure of the network (the number of interactions, the presence of hubs, clusters, network motifs, etc.).

The strategy to follow in order to establish whether a given network is stable is to mathematically analyze its structure. Ricard V. Solé and José M. Montoya (Solé & Montoya 2001; Montoya & Solé 2001) show that networks depicting certain patterns, like small world or scale free, because of their topological features, are quite robust to perturbations (mainly random removal of species). A network showing a small world topology is a network in which only few steps are needed to connect one node with another, while a scale free topology is a network with few hubs (nodes

[16] The distinction between a graph and a network is not straightforward. Matematicians usually talk about graphs, with their vertex and edges, while when we speak about real systems, like biological food webs, ecological communities, etc, we talk about networks, with nodes and links.

with many interactions) and many nodes connected with few or just one other node. To illustrate this, let us introduce an example from ecology. Solé and Montoya (2001) want to explain why complex food webs[17] in general show a high robustness (one type of stability) under the presence of perturbations. They analyze three of the most studied food webs in ecology (Ythan Estuary, Silwood Park and Little Rock Lake), and realize that the network structure of the three of them follow a Small world topology. They run a series of simulations on the three webs by removing some species at random. The simulations show that the three webs depict a great source of homeostasis by providing fast responses to the perturbations (Solé & Montoya 2001, p. 2040). Drawing on other investigations (Watts & Strogatz 1998; Strogatz 2001), they claim that the stability upon perturbations showed by the food webs is due to its topological structure. Different networks displaying the same topological structure, they claim, like technological, metabolic and neural networks, show a similar response to perturbations (i.e. stability). Thus, the explanation of the stability of these different network structures is very similar in all of them, turning the mechanistic details irrelevant for the *explanandum* at hand.

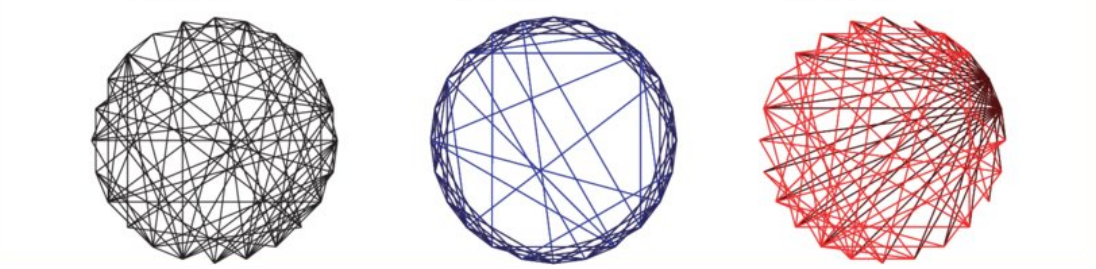

Figure 1. This figure exemplifies different types of networks. The one on the left is a random network. The one in the middle is a small world network. The one of the right is a scale free network. (From Sporns et al. 2004: 419).

To sum up, topological explanations, paradigmatically found in network models, appeal to the mathematical properties of a systems S' and then extend these properties to the real system S, that S' is supposed to represent[18].

There are examples of topological explanations in molecular biology too. A canonical example is given by Nicholaos Jones (2014). Jones claims the explanation of the vulnerability of the human immune system to attacks on the CD4+ T-cells (Kitano & Oda 2006) is due to the topological structure it deploys. Jones illustrates topological explanations by studying Hiroaki Kitano and Kanae Oda (2006) analysis of the aforementioned vulnerability. Kitano & Oda's strategy to explain the vulnerability of the human immune system to attacks on the CD4+ T-cells is to

[17] A food web is a graph representing different food chains. Usually a food web shows different plants, animals and other organisms, being connected by arrows representing a prey-predator interaction.

[18] However, not all network models are instances of topological explanations. There are cases in which networks are used mainly to represent. Thus, each case of network models needs to be evaluated, and what is more important, the nature of the *explanandum*, as Huneman suggested, will determine whether the role networks play is merely representational or explanatory.

characterize a biological pathway[19] for the adaptive human immune system and showing how the *explanandum* can be inferred from the pathway's architectural structure. The human immune system displays a bowtie structure in which CD4+ T-cells are non-redundant elements in the core of the network. Accordingly, CD4+ T-cells is a necessary step for every biological pathway of the adaptive immune system. If there is an attack to the CD4+ T-cells, no matter the mechanistic details of the attack (HIV, EVB i.e. mononucleosis…), the biological pathways responsible to block the attack will be defective. Jones points out that the explanation of this vulnerability appeals to a specific network structure of the adaptive immune system. Thus, the fragility of the system to attacks on that particular T-cells -located in the core of the network- is derived from the architectural structure it instantiates, i.e., form the mathematical properties of the network structure.

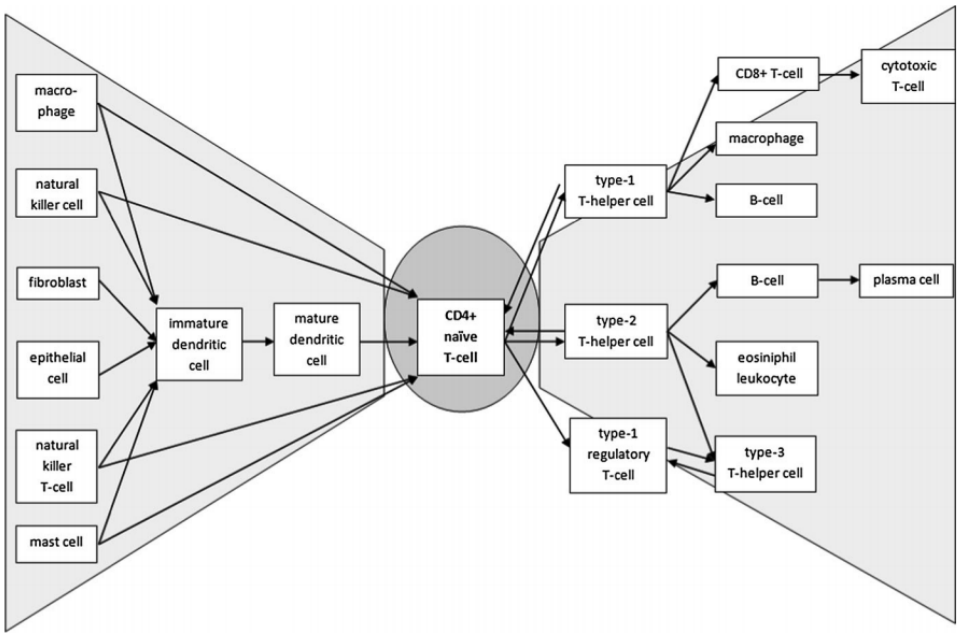

Figure 2. Bowtie structure of the immune system, with the CD4+ T-cells in the core of the bowtie. The mathematical properties of the network structure of the human immune system explains its vulnerability to attacks on CD4+ T-cells. (Figure taken from Jones 2014, p. 1138).

Other examples of topological explanation can be found in evolutionary biology (Huneman 2010) and systems biology (Green 2017, 2019).

The contrast between topological explanations and mechanistic explanations is the causal story. Mechanistic explanations are causal explanations. They are not complete until a causal story is told. Topological explanations on the contrary are non-causal, they explain by appealing to mathematical properties rather than by identifying causal information.

[19] Jones provides the following definition of a biological pathway: "a biological pathway is a directed network of interactions or functional relationships between components of a biological system operating in concert to accomplish a biological function in respons to environmental stimuli (Jones 2014, p. 1137)

Equilibrium explanations.

Cases of equilibrium explanations have been less discussed within the philosophical debates about scientific explanation. Huneman (2018c), as has already emerged previously, claims equilibrium explanation are a subtype of structural explanations, thereby maintaining that its explanatory power does not comes from capturing causal mechanistic information. However, Huneman does not examine the question of what makes equilibriums explanatory. In a similar vein, Rice (2015) shows that optimality models provide equilibrium explanations that do not appeal to information about the causal story of the *explanandum*, thus being non-causal explanations. Elliott Sober (1983) also advocates for a non-causal interpretation of equilibrium explanations. However, there are other philosophers who try to address the analysis of equilibrium explanations in a causal way. Angela Potochnik (2015) for instance, claims equilibrium explanations should be considered causal explanations and thus, addressed by causal accounts. However, she is aware that many causal accounts cannot accommodate such explanations, like the New Mechanism account. She provides her own way to accommodate them (Potochnik 2015). Part of Potochnki arguments are taken from Jaakko Kuorikoski (2007), who argues that we can give a causal interpretation of explanations appealing to equilibria.

The discussion around equilibrium explanations goes back to Sober (1983). In that paper, Sober writes against critics of Hempel which were claiming that all explanations must cite the causal story of the *explanandum* phenomenon. Sober's main contention is that there are explanations appealing to equilibrium states that do not use causal information in the *explanans*. An equilibrium explanation considers equilibrium states of a system, namely, a stable state that is maintained as a result of interacting forces in the system (Kuorikoski 2007). Usually, systems in equilibrium states have a certain domain of attractions, meaning that, if a system gets disrupted from its equilibrium, it can go back to his stable state. Examples of equilibrium explanations in biology can be found in ecology: for instance, we can ask why bacterial species in the human microbiome persist along the whole life of the host, resulting in an equilibrium state that maintains the host-symbiont system stable (Blaser and Kirschner 2007). Moreover, in evolutionary biology, equilibrium explanations features in explanations of organism's traits that appeal to the selective advantage conferred by the trait (Potochnik 2015, p. 1164), and in explanations that appeal to evolutionary game theory, like the equilibrium between two different behaviour strategies in a population (Smith & Price, 1973). The question at stake is why such states are explanatory, and whether they can be deemed mechanistic explanations or not.

Sober (1983) illustrates equilibrium explanations with Fisher's explanation of the 1:1 sex ratio in many species. Sober shows how Fisher's explanation of that phenomenon does not appeal to the causal trajectories that lead a population with certain proportion of males and females (initial conditions) end up in the 1:1 sex ratio. Instead of that, Fisher shows how regardless of the initial conditions, regardless of the evolutionary forces at stake, the final state of that population will have the same proportion of males and females. In fact, according to Fisher the equilibrium is the result of a difference in fitness for "there will be a reproductive advantage favoring parental pairs that overproduce the minority sex" (Sober 1983, p. 201). For instance, a causal explanation of a population's sex ratio in a given time t would appeal to previous initial conditions of the same population (the initial sex ratios) and the evolutionary forces that moved the population toward its final state. However, as previously claimed, this is not how Fisher explains the equilibrium, he actually claims that the selective forces at stake and the initial conditions do not matter at all, for whatever they are, the resulting state will be always the same equilibrium.

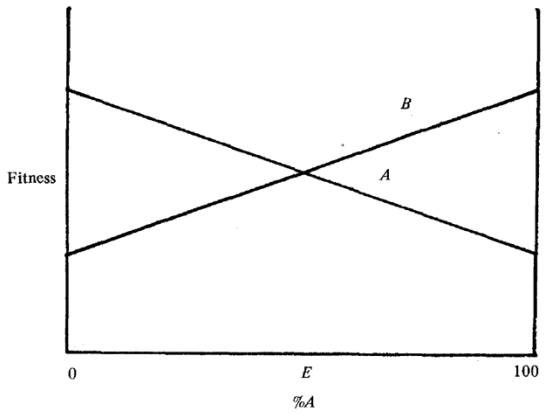

Figure 3. The diagram shows the fitness functions of traits A and B in a population. The function tells the fitness of each type (in expected number of offspring) given the frequency in the population. *E* is an equilibrium point in which natural selection does not acts. All the other places, natural selection always advantages the minority trait and leads to the equilibrium value *E*.

Having clarified this, I should move towards the view that argues that equilibrium explanations have a causal interpretation. Potochnik (2015) for instance, distinguishes between explanations that track the causal story of an *explanandum* and those explanations that track causal patterns, as patterns that identify causal dependencies and the scope of those dependencies, namely the conditions upon which the dependence hinges. Potochnik argues that equilibrium explanations belong to the later form of explanation. Despite the fact that equilibrium explanations do not provide information about the causal processes that end up in the *explanandum* event, they are regarded as causal, for they provide information about certain causal dependencies and their scope. Potochnik (2015) takes this idea from Kuorikoski (2007). Replying to Sober (1983), Kuorikoski claims that equilibrium explanations provide information about the relationship between structural features of the system and the phenomenon to be explained. Appealing to Woodward's interventionist account of causation (2003), Kuorikoski claims that such explanation does not capture causal information as causal processes, but it captures information about a causal dependency. You could change a structural property of the system that would change the equilibrium point, although, the relationship between the structural features and the equilibrium value is invariant under such interventions. Potochnik (2015) illustrates this point with a toy example, explaining the temperature of the coffee within one's cup. The explanation of the equilibrium state that the coffee reaches (temperature), will not appeal to the previous temperature of the coffee and how the molecules of the liquid substance behaved so as to reach the given temperature. The explanation, Potochnik contends, would appeal to the actual temperature of the room in which the cup of coffee is, and the fact that the cup has been there for many hours. If we were to change the temperature of the room, the equilibrium value would change. And yet, the dependency between room temperature and coffee temperature would still be invariant (Potochnik 2015).

As appealing as this strategy may seem, it reveals a limitation of the New Mechanism account of explanation. As Potochnik (2015) argues, those explanations, even if deemed causal, do not provide causal-mechanistic information about the causal story of the *explanandum*. This said, the question of whether equilibrium explanations are causal in this other way suggested by Potochnik and Kuorikoski is still object of debate.

I turn now to illustrate the same kind of problem by focusing on evolutionary game theory. Evolutionary game theory is a discipline where the use of equilibrium explanations is pervasive. John Maynard Smith and George R. Price (1973) for instance, showed how to apply game theory to analyze and therefore explain the behaviour of animal conflict, or to be more precise, the proportion of different behaviours in a population. To illustrate briefly, Smith and Price depict scenarios of animal conflict as a two-player game in which each player can apply a particular strategy. Their model describes three ways to behave in a conflict: D, attacks that likely seriously damage its opponent; C, attacks unlikely to cause seriously injury to its opponent; R, retreat. They define 5 strategies (Mouse, Hawks, Bully, Retaliator, Prober-Retaliator), as "a set [sets] of rules which ascribes probabilities to the C, D and R plays, as functions of what has previously happened in the course of the current contest" (Smith & Price 1973, p. 16). For instance, the strategy "Mouse" goes as follows: "Never plays D. If receives D, retreats (…). Otherwise plays C until the contest has lasted a preassigned number of moves" (Smith & Price 1973, p. 16).

They simulate 2000 contests for each of the 15 possible scenarios with what they call pseudo random numbers given by an algorithm that varies the contests; for the simulation, they ascribe different probabilities like the following: probability that D creates a serious injury in one attack is 0.10, and calculate the pay-off matrix for the simulation, as "measures of the contribution the contest has made to the reproductive success of the individual" (Smith & Price 1973, p. 15).

With the pay-off matrix they introduce the fundamental notion of an Evolutionarily Stable Strategy (ESS henceforth). An ESS is a strategy that when played by most of the members of a population will be stable under natural selection, namely, that no other rival strategy could displace it, or in other words, no other strategy would confer to the individuals higher reproductive success, higher fitness. All ESS are cases of Nash equilibria (Nash 1950), which, as defined in game theory, is a strategy in a two-player game stating that no player would get a higher pay-off by changing strategy unilaterally.

Eventually, Smith and Price (1973) show that Retaliator is an ESS, for the pay-off values of the other 4 strategies when played against it are lower that the value Retaliator gets when playing with himself. That means that in a population where most of the individuals play Retaliator strategy, the mutant appearance of one of the other four could not displace it, for they would be mainly fighting/playing against Retaliators thus getting lower pay-off than its opponent[20].

What interests us here is to focus on how Smith and Price explain the fact that in a population in which there is conflict, there will be an equilibrium point where most of the individuals play Retaliator strategy and few other will be playing the other 4 strategies, no matter the initial conditions, say the initial proportions of strategies played. The appeal to an equilibrium point, given by the notion of an ESS, is what is playing the explanatory role and not the identification of causal trajectories of particular populations, analogous to the example given in Fisher's sex ratio.

It is therefore clear that the New Mechanism account of scientific explanation is not fruitful in analyzing how explanations work in evolutionary explanations that appeal to game theory.

[20] Obviously, their models are idealized for they omit several details and make their models depend on certain assumptions which are definitely false. However, their models provide good explanations and good predictions about the proportions of behaviour in animal conflict.

		Opponent				
		"Mouse"	"Hawk"	"Bully"	"Retaliator"	"Prober-Retaliator"
Contestant receiving the pay-off	"Mouse"	29.0	19.5	19.5	29.0	17.2
	"Hawk"	80.0	−19.5	74.6	−18.1	−18.9
	"Bully"	80.0	4.9	41.5	11.9	11.2
	"Retaliator"	29.0	−22.3	57.1	29.0	23.1
	"Prober-Retaliator"	56.7	−20.1	59.4	26.9	21.9

Figure 4. Pay-offs matrix of the five different strategies of animal conflict (taken from Smith & Price 1973, p. 16).

1.3.2. A neo-Hempelian covering law account of scientific explanation

A second strategy that put some limits to the wide scope of mechanistic explanation, is the one that points towards the claim that laws of nature still play a role in scientific explanation, and in particular in biology. This strategy was very popular due to Hempel's DN model and IS model, though it is under discussion nowadays. In the domain of biology in particular, it has been claimed that there are not laws of nature, at least as laws have been traditionally understood (Smart 1963; Beatty 1995; Rosenberg 2001). However, several authors argue that talking about laws in biology is not only possible but necessary (Carrier 1995; Brandon 1997; Mitchell 2003; Lorenzano 2007; Leuridant 2010; Dorato 2012; Díez and Lorenzano 2013, 2015; Alleva et al. 2017). Following this strand of thought, Díez provides a neo-Hempelian account of scientific explanation, which is entitled ASE (ampliative specialized embedding), that regards scientific explanations as covering law in a contemporary way (Díez 2014; Alleva et al. 2017). ASE maintains that some explanations work by showing how an *explanandum* is expectable under the existence of lawful generalizations, even in biology. Before introducing ASE, it's important to dwell on the question of whether there are laws of nature in biology for a moment, as without this possibility it would be difficult to argue in favour of a covering law model. Sandra Mitchell's account of pragmatic laws (Mitchell 2000) is a promising way to understand laws in biology nowadays.

Pragmatic laws

Sandra Mitchell account of pragmatic laws is one of the most promising strategies to keep the importance of the notion of law of nature in biology (Mitchell 1997, 2000). Mitchel's proposal intends to change the framework from which philosophers of science used to investigate the notion of law. She aims to avoid the question "what we should call a law" and aims at achieving "an understanding of how scientific generalizations of various types function in inference to satisfy the pragmatic goals of science [predicting, explaining and intervening]" (1997: S469). Mitchell is aware of the failures committed by those who tried to identify necessary and sufficient conditions for a generalization to be a law (what she dubs "the normative approach"). Instead, she proposes a pragmatic approach that rather than focusing on necessary and sufficient conditions, looks at the conditions of applicability of such scientific claims.

The normative approach offers an analysis of laws of nature based on a twofold distinction: on the one hand, *the lawful vs accidental* distinction, on the other hand, the *necessary vs contingency* distinction. Proponents of the account depart from a definition in terms of necessary and sufficient conditions and then check whether the candidate laws fulfil the criteria. Thus, laws of nature have four essential features according to the normative position (Mitchell 2000):

- logical contingency (have empirical content)
- universality
- truth
- natural necessity.

Normative accounts tend to “mirror natural necessity with logical necessity”, and this is something that makes the property of lawhood being an “all or nothing property”, black or white. Either a generalization is nomological or it is not, there is no place for intermediate stages. According to Mitchell (2000), such view does not allow to take as laws some generalizations from science that capture "strong causal relation between events" without fulfilling the stronger conditions of nomological necessity, i.e. they are not generalizations without exceptions, think for instance about Mendel’s laws which do suffer from exceptions (Lorenzano 2007). It looks therefore that such philosophical analysis of laws leads to a dead-end road, because only few, if any, principles not only in biology but even in physics fulfil such stringent conditions (Cf. Dorato 2012). Besides, the notion of natural necessity is extremely difficult to properly define and explain, and different metaphysical analysis thereof are not compelling (Dretske 1977, Armstrong 1983).
Following this normative strategy, let’s focus on John Beatty’s Evolutionary Contingency Thesis (Beatty 1995) in which he claims that "all distinctively biological generalizations describe evolutionarily contingent states of nature" (Beatty 1995, p, 217), thus there are no laws in biology because laws, Beatty claims, are supposed to be more than just contingently true. According to Beatty this contingency comes in two different varieties, a weak one and a strong one. Weak contingency points towards the fact that generalizations in biology are always context dependent, namely what makes a trait to be advantageous for an individual, is always dependent on external conditions. Those conditions being changed, the trait might lose its advantageous status. Strong contingency is illustrated by the famous tape example from Stephen J. Gould (1989). Gould claimed that if we could restart the life on Earth from the beginning, having the same initial conditions and the same selective forces the results of evolution would be radically different than they are today[21].

Despite the contingent nature of biological generalizations, Mitchell (2000) claims that all generalizations from science suffer from the same two types of contingency. On the one hand, all generalizations are logically contingent, i.e. their truth does not depend on a logical relation, but on certain empirical facts. On the other hand, Mitchell claims that all generalizations are "evolved" inasmuch as the relations they describe are all dependent upon certain conditions obtaining. For instance, in order for Galileo's free fall law to describe real states of affairs, the Earth need to have the mass it has, "if the core of the Earth were lead instead of iron, the quantitative acceleration would be four times what it is" (Mitchell 2000, p. 251). Therefore, according to Mitchell, the truth of every generalization depends on certain conditions obtaining, like Galileo's free fall law depends upon the structure of the core of the Earth.
Nonetheless, Mitchell points out, Beatty is right to claim that there seems to be a difference between Mendel's law of segregation and the law of free fall (the former is much more domain

[21] In the same vein, Lorenzano (2007) claims that contingency is mainly due to random mutation, functional equivalence (the fact that that there are different ways to adapt to a change in the environment), and random genetic drift, mainly acting in small populations.

restricted and with exceptions while the other is more widely applicable). Mitchell believes that such distinction cannot be properly captured within the normative framework by deploying the contingency vs necessity dichotomy. That framework ascribes many generalizations to the realm of accidents, thereby failing to give the attention they deserve given their ability of fulfilling the pragmatic goals of science.

What is really needed in order to capture the distinction between Mendel's law of segregation and Galileo's free fall law is a new framework. This new framework, Mitchell contends, allow us to ”locate different degrees of stability of the conditions upon which the relation described is contingent[ly true]" (Mitchell 2000, p. 252). Therefore, the normative strategy of framing natural necessity/contingency in the same terms as logical necessity/contingency generates a misleading answer to the question about the difference between generalizations in physics and in biology. According to Mitchell there is no such a clear-cut *qualitative* difference tracked by the lawful vs accidental distinction. Rather, the difference between generalizations in physics and in biology is a matter of degrees: the conditions that make certain generalization in physics to be true are more stable that the ones that make biological generalizations true. For instance, the stability of the conditions that make Galileo's free fall law to hold are stronger than the ones that make Mendel's law true. The acceleration of a body (as stated by Galileo's law) under those conditions, is deterministic, while 50-50 segregation in sexual organisms under the conditions that make such law true, are probabilistic. Therefore, there is a difference of stability and of strength.

Thus, Mitchell’s pragmatic notion of laws focuses on the use of a scientific generalization regarding the pragmatic goals of science: predicting, explaining and intervening. Philosophers of science should wonder "how do they [scientific generalizations] function to allow us to make predictions, explanations and successful interventions" (Mitchell 2000, p. 259). So, every generalization that contributes to the pragmatic goals of science will be attributed a certain degree of nomological force. This nomological force will depend on what Mitchell calls the characteristic space of each generalization, namely, how a system must look for a generalization to be successfully applicable.
This minimum characterization is enough to keep the notion of law still at play in biology. According to this pragmatic account, those generalizations (or models) that allow to explain, predict and intervene are attributed a given nomological force, and can then be considered scientific laws. Moreover, the idea that nomicity is also present in biology as not merely accidental generalizations with counterfactual import, no matter how *ceteris paribus,* local or domain restricted, is also defended by Díez (2014) in his neo-Hempelian account.

Neo-Hempelian account

Taking this wide notion of law into account, let us now turn to canvass a very recent covering law model of scientific explanation entitled ASE (Díez 2014, ASE standing for Ampliative Specialized Embedding). This model is highly inspired by Hempel’s ideas about scientific explanation although it is in a better position to deal with contemporary issues in explanation.

ASE’s account of scientific explanation has its roots in the idea that to explain a phenomenon is to embed it into a nomic pattern within a theory net, that is to say, within a hierarchical and synchronic structure of a theory (Díez 2014). This account should be framed within the structuralist tradition in the philosophy of science (Balzer et. al 1987; Díez & Lorenzano 2002).

The *structuralist* tradition has as its main goal the formalization of scientific theories following a model-theoretic account (see Díez & Lorenzano 2002, introduction, for details).
ASE weakens Hempel's model on the one hand and strength it on the other hand. First, ASE contends that the relation between *explanans* and *explanandum* is an embedding between models rather than a deductive and inductive inferential relation between statements. By weakening this relation, ASE avoids the non-necessity problems (Mayor with paresis) Hempel's original characterization had to face, for it allows expectability with low probability, i.e. the explanans rises the probability of the *explanandum*. In ASE, the *explanandum* is represented as a *data model*, containing certain entities and functions defined on them, while the *explanans* is represented as a *theoretical model*, defined using certain laws and containing at least the same kind of entities and concepts as the data model. Second, Díez, strengthens Hempel's account by claiming that embedding relation between models has to be ampliative and specialized in order to count as a scientific explanation:

- *Ampliative*: The *explanans* (theoretical model) has to use t-theoretical concepts, namely some new conceptual/ontological machinery not present in the *explanandum*. Thus, the *explanandum* has to be defined using T non-theoretical concepts. For instance, space and time are T-non theoretical for classical mechanics while force and mass are T-theoretical. That means that in order to measure forces and masses we need to presuppose always a mechanical law while with space and time we do not have to. This distinction is close to Lewis' on "new" and "old" vocabulary and Hempel's on "characteristic'' and ''antecedently understood" terms (Díez 2014). Moreover, a concept can be T-theoretical in a theory and T-non-theoretical in another. "Pressure" for instance is T-theoretical in classical mechanics and T-non-theoretical in thermodynamics.
- *Specialized*: The *explanans* has to make use of a special law, a specialization of a guiding principle of a theory, in order to avoid illegitimate or *ad hoc* embeddings. As Díez puts it, this condition demands that "the specific regularities used in the specific embedding must be non-ad hoc, acceptable relative to the current state of the framework, either because they already exist (as accepted) or because, if new, they are acceptable compared to the regularities that already exist" (Díez 2014, p. 1438). This condition is well exemplified by those highly unified theories that follow a theory-net structure, with a hierarchy structure having a guiding principle on the top and special laws that specialize this principle in its ramified branches.

So, a model or theory T explains the phenomenon P iff (i) P is nomologically embeddable in T, (ii) T includes T-theoretical concepts and (iii) T makes use of special laws.

A data model, an *explanandum*, will be thus expectable given the theoretical model. For instance, according to Díez, in the case of classical genetics, "the *explanandum* is the data model that describes certain transmission of phenotypes, and the *explanans* is the theoretical model that includes genes and is defined by certain genetic laws. We explain the transmission of traits if we succeed in embedding the data model into the theoretical one, that is, if we obtain the observed phenotype sequence from the genetic model" (Díez 2014, p.1420).

One of the possible criticisms made to ASE is that the account can only be applied to analyze the nature of scientific explanations concerning those theories that are structured in a theory net-hierarchical way, that is, theories that are highly unified. This would be a problem insofar as there seem to be several theories in biology that do not have such structure. Besides, nowadays biologists speak and create models, replacing the notion of theory as the main element of analysis.

Even if this might pose a problem for the monistic aspirations of ASE's account, Díez himself claims that it is not necessary to formalize a theory or model in a model-theoretic way in order to analyze the nature of scientific explanation following his account. Even if there are theories or models, just like the ones New Mechanists analyze in biology or neuroscience, that do not have such a theory-net structure, we can still use ASE to analyze their explanatory power. ASE contends that we can identify some type of generalization (domain-restricted and non-exceptionless, if that's the case) involved in the construction of some biological models, as well as the appeal to new conceptual/ontological machinery with respect to the explanans. These features, being specific enough and non-*ad hoc*, make the explanandum somehow expectable. Those are, according to Díez, the features that allow us to speak about a covering law model.
Furthermore, ASE highlights explanatory strategies that do not seem to be mechanistic, for they do not appeal to causal constitutive details but to generalizations as the main posit of explanations (see Alleva et al. 2017 for a specific analysis of allosterism following this lines).

It must also be noted that, to overcome the possible criticism to ASE against the existence of laws in biology, ASE can always appeal to Mitchell's notion of pragmatic laws. Thus, we can add this account to the pluralist picture of explanation in biology described so far. This covering law account would complement causal-mechanistic analysis and mathematical non-causal analysis of explanation in biology[22].

1.4. Explanatory pluralism in biology

Section 1.1 depicted a scenario in which, given the problems that the traditional accounts of scientific explanation faced - e.g. the fact that none of them manage to provide a compelling analysis that captures how scientific explanation works in domains as different as biology, physics, neuroscience, psychology etc. - two possible strategies could be followed. On the one hand, one can pursue the search of an even better general and monist account of scientific explanation (whose research would follow questions 1) to 3) of section 1.1. On the other hand, one can advocate for a *pluralist* strategy. The discussion pursued so far shows that even if the notion of mechanistic explanation is successful in capturing the explanatory power of a wide range of biological models, there are several occasions in which the explanatory force of a model is not given by the appeal to causal mechanistic information. This suggests some kind of explanatory pluralism.

Surely, as urged by Braillard and Malaterre (2015), a pluralist approach to scientific explanation must take up a number of questions, I will consider two of them: i) are there good reasons to endorse explanatory pluralism besides the fact that no monist account has been found? ii) Do different explanatory accounts compete with one another or are rather complementary? In case they are complementary, do different explanatory accounts target different types of explananda, i.e. belonging to specific fields or scientific disciplines?
By replaying to questions i) and ii), this section will introduce the notion of *weak explanatory pluralism*, whose key contention is that different explanatory accounts are complementary rather

[22] Whether ASE captures universal features of scientific explanation is something controverted. Díez claims his account is minimal enough to be applied everywhere. I am rather cautious in doing such a claim for I believe more needs to be made in order to show that ASE can add something to the analysis of explanations carried on by the New Mechanism philosophy.

than competing. Furthermore, I will briefly introduce the notion of *integrative pluralism* as a strategy that captures how a complex problem in biology must be addressed by multidisciplinary research and therefore be explained by the integration of two or more disciplines or models.

Question i), asking whether there are independent reasons for preferring explanatory pluralism, as opposed to explanatory monism, must be related to the more general question of the unity or disunity of science, or even more general about scientific pluralism vs scientific monism. Scientific monism is the claim that "the ultimate aim of science is to stablish a single, complete and comprehensive account of the natural world (…) based on a single set of fundamental principles" (Kellert et. al 2006, p. x). A scientific monist would claim that theories and models should be evaluated on the basis of their ability to provide a complete account based on fundamental principles. She might admit that a plurality of models and approaches is used by different disciplines and account for this fact by contending that this is due to contemporary science's incompleteness. A scientific monist like Kitcher (1993) might even claim that this plurality of competing models and methods is actually important for scientific progress but that, eventually, we must reach a unifying picture of science. A scientific pluralist though, claims that this plurality of models and approaches does not reveal anything about the maturity of a science, but it is rather correlated with the complexity of that science (Mitchell 2002), and by no means constitutes a deficiency. Furthermore, a pluralist claims that not all phenomena in nature can be completely and comprehensibly explained and represented using a single set of fundamental principles. Therefore, the diversity of methods and approaches applied to explain and represent such phenomena should not be considered a feature of an immature science (Kellert et. al 2006). Thus, a scientific monist would advocate for the search of an even better monist account of scientific explanation. A scientific pluralist on the other hand would accept that science needs to appeal to a plethora of different approaches and explanatory accounts to explain complex phenomena in nature[23].

Focusing on question ii) about the relation between the different explanatory accounts, Brigandt (2013a), claims that at least two different accounts of explanation complement each other in biology. On the one hand, in ecology and evolutionary biology, explanations appeal to quantitative generalizations, namely, they use mathematical models to represent and explain in a substantial way. More specifically, mathematics is used to represent and explain the dynamics of biological systems. For instance, as has emerged in section 1.3, network models are used to explain different features of ecological communities like stability, robustness…, and how they vary in time; population genetics is used to explain change in gene frequencies, or the size and structure of populations. Of course, not all explanations in ecology or evolutionary biology are of this form, for we could find as well causal mechanistic explanations there, if we formulate the adequate questions, or target specific *explananda*. In Brigandt words:

> A mathematical model from population genetics may predict an increase in the prevalence of a phenotype given that its fitness is larger than the fitness of other phenotypes. An application to a concrete situation has to causally explain why their specific phenotypic traits endow some organisms with a reproductive success (fitness) higher than that of organisms possessing other

[23] Later, Kitcher (1999), who has extensively argued for unification in science, acknowledge that unification in biology is impossible due to the complexity of the living world. However he believes that unification should be a regulative idea, claiming that biologists should try to achieve as many unification as nature allows.

traits, based on the organisms' actual environmental conditions, i.e., a causal explanation of why natural selection favors some phenotypic traits over others in this context.

On the other hand, Brigandt (2013a) claims that in experimental biology (physiology, molecular biology, developmental biology, systems biology…) explanations are mainly causal mechanistic. Nevertheless, as this introduction has already shown, there are explanations in such experimental fields that are not causal mechanistic, but rather mathematical. In domains such as systems biology or developmental biology, we can as well find mathematical models that not only describe the dynamics of systems but explain some features of those systems (Brigandt 2010, 2015). According to Brigandt (2013a), this picture implies a *weak explanatory pluralism*: there are biological phenomena that are explained via mathematical explanations, there are others explained by causal mechanistic explanations and, as is suggested in the present introduction, by covering law explanations as well. These explanations are not competing for they apply to different phenomena. This weak pluralism contrasts with a strong pluralism defending that there might be competing explanations for the same phenomenon. Strong pluralism seems to be an untenable position, for it involves accepting some sort of contradiction, say, the existence of two valid and incompatible explanations for the same phenomenon. A more charitable reading though would be to understand this strong pluralism, rather than contradictory, as illustrating some kind of scientific relativism.

Reading through the contemporary literature in philosophy of biology reveals that *weak explanatory pluralism* seems to be the most popular approach to scientific explanation in biology (Mitchell 2003; Brigandt 2010, 2013a; Braillard & Malaterre 2015).

Moreover, philosophers like Sandra Mitchell (2003), Alan C. Love (2008), Ingo Brigandt (2013a), Constantinos Mekios (2015), Maureen O'Malley and Orkun S. Soyer (2012), and Sara Green (2017, 2019), to mention just a few, acknowledge the fact that many complex phenomena in biology, like in systems biology, developmental biology or evolutionary biology, require explanations that usually appeal to more than one discipline. They use the term "integrative pluralism" (Mitchell 2002, 2003) to describe this phenomenon. Integrative pluralism is an anti-reductionist stance, as it denies that some disciplines can and should be reduced to others for explanatory purposes, like classical genetics to cell biology[24]. However, an anti-reductionist stance does not imply that different disciplines are completely autonomous. In fact, the notion of integrative pluralism shows how some complex problems cannot be solved just by one discipline, creating an epistemic framework of multidisciplinary research. To talk about multidisciplinary research and thus integration between two or more fields, Love (2008) introduces the notion of a *problem agenda*. A problem agenda is a complex problem which requires the resolution of a set of questions. By replying to these questions, one produces an adequate explanation. Each component question has some criteria of explanatory adequacy which establishes what counts as a good explanation. According to Love (2008), the criteria of explanatory adequacy of a problem agenda determines which are the disciplines needed to solve the problem, or in other words, to provide adequate explanations. For instance, explaining the origin of novelties in evolutionary biology could be seen as a problem agenda, involving different disciplines like paleontology, phylogeny, developmental biology… To explain the origin of evolutionary novelties Brigandt claims that, "a first step would be to lay out a sequence of relevant morphological traits in different species within a phylogenetic lineage, leading up to the presence of the novelty" (2013, p.82). To

[24] Reductionism is mainly an account of the relation between scientific theories or disciplines.

do so, paleontology and phylogeny are essential, the former provides data about the fossil record of the structure under study, whereas the latter explains at which junctures of the phylogenetic tree the morphological changes took place. Developmental biology is important as well, for the fact that an ancestral organism developed so as to end up having the new structure (new morphological trait) stands in need of an explanation (Love 2008). Thus, to explain the origin of novelties in evolution, many different disciplines are needed as well as an integration between the models provided.

In the same vein, Mitchell (2002, 2003) argues that integrative pluralism provides an image of compatibility of models and explanations in science. She contends that sometimes in biology we find the presence of different models tackling parts of a bigger problem (or problem agenda in Love terms). For instance, in explaining how division of labour works in insects, biologists are engaged in two main lines of research, generating different models that focus on specific features of labour in a community (Mitchell 2002). There are adaptive explanations on the one hand, appealing to optimality models analyzing which patterns would be optimal in the division of labour in terms of "ergonomic efficiency". If there were communities instantiating in the past different patterns than the optimal one ("age-related organization with specialization[25]", Mitchell 2002), they would have disappeared in the struggle for life. On the other hand, Mitchell states that adaptive explanations "black boxes" the mechanisms that lead to the generation of the pattern, thereby neglecting self-organization models as important elements for the explanation of the division of labour. In Mitchell's words, "[adaptive explanations] ignores the physiology or development of the colony phenotype. (…) However, how a trait develops can and does restrict the range of adaptive explanations that could be plausible entertained" (Mitchell 2002, p. 59). According to Mitchell, a self-organization model shows "how complex collective behavior can emerge from interactions of individuals exhibiting only simple behaviors without the need for a central organizing agent" (Mitchell 2002, p. 59). However, at the same time, there seems to be three available and theoretically incompatible self-organization models, the three of them though capturing and explaining partial information about division of labour. They are incompatible, Mitchell claims, for they refer to different idealized systems. On these grounds, Mitchell (2002) wonders how such different and *a priori* incompatible types of models need to be treated in order to provide a complete explanation of the division of labour. She claims that even if they are incompatible at the theoretical level, when applied to a concrete community the conditions that the three models describe in isolation obtain in the system analyzed; so, each of them does provide partial explanations of the complex problem. Going into the details of how this integration works goes beyond the scope of this introduction. The take-home message for our discussion is this: while dealing with complex problems in biology, you might need to integrate different models that, due to idealized requirements, might be incompatible. Such models might even provide explanations of different kinds; however, given a suitable integrative framework, such models are all part of the explanation required.

[25] An "age-related organization with specialization pattern" in the division of labour of a community is a pattern that depicts the following features: age-polyethism, which refers to the regular changes in the tasks performed by an individual in the community along his life cycle; homeostatic regulation, which refers to the ability of the community to adapt the proportions of members working in each task due to internal and external conditions; Individual specialization, describing the ability of a single individual to adjust his task in order to maintain homeostasis in the community.

1.5. Concluding Remarks

Along the lines of this introduction, I have briefly shown that the monist aspirations of traditional accounts of scientific explanations were at odds with actual biological explanations, suggesting a pluralistic interpretation of scientific explanation in this domain.

I have presented the New Mechanism account of scientific explanation, i.e. mechanistic explanation, which is the dominant account to analyze scientific explanation in biology nowadays. I have shown that despite the high success of mechanistic explanation in capturing the workings of explanation in biological models, its scope is still something controverted. I have opened two lines of criticism that limit the scope of mechanistic explanation. These two lines provide examples in which biological explanations seem to gain its explanatory power not by identifying a mechanism with its causal stories. Mathematical explanations (topological and equilibrium), and ASE, a covering law accounts, limit the scope of mechanistic explanations even in models that could be thought to provide mechanistic explanations.

Finally, I have briefly analyzed what type of explanatory pluralism is nowadays widespread within the philosophy of biology, mentioning the stance of a *weak explanatory pluralism* and the notion of *integrative pluralism*.

2. AIMS OF THE RESEARCH PAPERS AND RESEARCH METHODOLOGY

2.1. The papers

The three papers of the present thesis dissertation fall within the two lines of criticism to mechanistic explanation advanced in section 1.3 of the introduction. In what follows I am briefly presenting, in a schematic way, the main goals of each of them.

The first two papers, "Explaining the behaviour of random ecological networks: the stability of the microbiome as a case of integrative pluralism" (Deulofeu et al. 2019) and "Equilibrium explanations as structural non-mechanistic explanations: the case of long term bacterial persistence in human host" (Suárez & Deulofeu Forthcoming), fall within the first line of criticism advanced in section 1.3. Both papers exemplify biological models that use mathematical modelling to explain (and not only describe) in a non-mechanistic way.

The third paper, "When mechanisms are not enough: the origins of eukaryotes and scientific explanation" (Deulofeu & Suárez 2018), fall within the second strategy mentioned in section 1.3. The paper analyses the explanation of one of the major transitions in evolution, the origin of eukaryotic cells, the transition from prokaryotes to eukaryotes. We apply a neo-Hempelian model of explanation to illustrate that the explanatory force of the endosymbiosis or exogenous theories of the origin of eukaryotes is not due to the description of a mechanism with its causal story but to two different posits: a pragmatic law and an ampliative theoretical novelty.

2.1.1. Explaining the behaviour of random ecological networks: The stability of the microbiome as a case of integrative pluralism

The paper "Explaining the behaviour of random ecological networks: the stability of the microbiome as a case of integrative pluralism" illustrates an extended version of a topological explanation. The paper has three main goals.

- The first goal is to provide a philosophical analysis of Coyte et. al (2015) biological model which applies network modelling to study and explain the stability behaviour of the human gut microbiome.
- The second goal is twofold. One to one hand the goal is to argue that rather than merely describing the workings of a complex system, the model provides an explanation of the stability behaviour of the system by appealing mainly to its topological structure (network analysis) and the types of interactions between the species conforming the microbiome (competition, cooperation…). On the other hand, the goal is to show that the model explains the stability behaviour not by identifying elements of the causal structure of a mechanism, even if a mechanism can be identified, but by providing a topological explanation complemented with the dynamics of the system, in terms of the proportion of the interaction types (competitive, cooperative and exploitative).
- The final goal is to show that the explanation falls within the scope of integrative pluralism, in which an explanation of a complex system behaviour needs to bring together knowledge from different disciplines, ecology and mathematics (network modelling) in this case.

2.1.2. Equilibrium explanations as structural non-mechanistic explanations: The case of long-term bacterial persistence in human hosts

The paper "Equilibrium explanations as structural non-mechanistic explanations: the case of long-term bacterial persistence in human host" illustrates a case of an equilibrium explanation. The paper has three main goals:

- The first goal is to provide a philosophical analysis of the explanation given by Blaser and Kirschner's nested equilibrium model of the stability of persistent long terms human-microbe associations.
- The second goal is to show that the explanation given by the model is not mechanistic in terms of describing the internal working of a system by its entities, activities and its causal interactions and organization. Rather, the explanatory force, we claim, is given by the appeal to a set of differential equations that together show how the system remains in a Nash equilibrium according to an Evolutionarily Stable Strategy. The nested nature of the model is a basic element to show how the evolutionarily stable strategy works, maintaining the equilibrium under the existence of perturbations.
- The third goal is to show that the explanation given by the model fulfils the standards given by Huneman as a structural explanation, in particular an equilibrium explanation.

2.1.3. When mechanisms are not enough: the origins of eukaryotes and scientific explanation

The paper "When mechanisms are not enough: the origins of eukaryotes and scientific explanation" intends to appeal to a neo-Hempelian account of explanation in order to show that the general explanation biology provides of the origin of the eukaryotic cell is not causal mechanistic but rather covering law, using a new conception of pragmatic laws given by Sandra Mitchell. The paper has four main goals:

- The first goal is to provide an analysis of the general explanation given by exogenous or endosymbiosis theories to the origins of the eukaryotic cell. We claim that even if there are different exogenous accounts of the origin of eukaryotes, all of them have symbiogenesis as the cornerstone.
- The second goal is to show that the set of questions an endosymbiosis theory of the origin of the eukaryotic cell must reply, are not mechanistic in nature. We show that even if you collect detailed causal information about the process (a requirement of mechanistic explanation), it does not give us the answer we seek.
- The third goal is to show that symbiogenesis can be seen as a pattern that allow biologists to make predictions, provide explanations and allow manipulations. According to these pragmatic goals of science, our aim is to show that symbiogenesis can be considered a pragmatic law in terms of Sandra Mitchell's account of pragmatic laws.
- The final goal is to show that a neo-Hempelian account of explanation is successful in identifying the explanatory force of the explanation biologists provide of the origin of eukaryotic cell.

Thus, these papers provide instances of explanations in biology that seem not to be mechanistic, or at least not fully mechanistic. Mathematics in some occasions and pragmatic laws in others provide the explanatory force of the aforementioned biological models.

2.2. Research methodology

The research methodology used in this thesis dissertation is the one of philosophical analysis, which in the analytic tradition is mainly conceptual analysis combined with case studies. On the one hand we did a wide revision of the literature of different positions regarding explanation in biology, mainly the New Mechanist positions and its critics. We made a critical analysis of those positions, providing comments about their strengths and shortcomings. We made some research looking for appropriate biological models to make our analysis of scientific explanation, and choose three biological models, taken from ecology, molecular biology and evolutionary biology, providing explanations of some biological phenomena. Finally, we made our own contribution to the debate, by claiming that the explanatory force of the three biological models do not rely on what the standards of mechanistic explanation consider.

The three papers thus have a similar structure. A literature review, where the different positions are presented, and the philosophical debates around scientific explanation in biology are introduced. An analysis of a case study, a biological model. A philosophical discussion about our own position contrasting it with others and contributing to the discussion framed in the first part.

3. CONCLUSIONS

In this work, I aimed to clarify the debate about the scope and limits of the New Mechanism account of scientific explanations in biology. The widespread and very successful use of mechanistic explanation in many areas of experimental biology in the last twenty years has led some mechanistic philosophers to think about the possibility of extending the account towards other branches of biology. However, this has required some revisions to important changes in the notion of mechanistic explanation, generating controversy on whether, after all, this account still makes the analyzed explanations mechanistic in kind.

The results of the three research papers that constitute this dissertation suggest that the New Mechanism account of scientific explanation is not the only successful account whereby we can capture the explanatory strategies of biological models. The first two papers (Deulofeu et al. 2019; Suárez & Deulofeu forthcoming) show that there are cases in which the role that mathematics plays in biological theorizing is not only representational, by describing biological systems and showing its dynamics, but explanatory *per se.* These two papers accomplish this task by analyzing the explanatory force of two biological models and arguing for the following two points: first, the models do not explain by capturing causal mechanical information of the phenomena under study; secondly, the explanatory force is given by the use of formal mathematical tools. In the first paper (Deulofeu et al. 2019) we have claimed the model gains its explanatory force by appealing to topological properties of the human gut microbiome, which is supposed to have a network structure. In the second paper (Suárez & Deulofeu forthcoming) the model gains its explanatory force by appealing to an equilibrium and nested state of the system (host with its symbionts), achieved by an evolutionarily stable strategy. These models, we contend, provide examples of structural explanations in biology following Huneman (2010, 2018a).

In the third paper (Deulofeu & Suárez 2018) it has been argued that laws of nature still play an explanatory role in biology, always under the re-definition of laws in terms of the pragmatics of science. A covering law account of scientific explanation, ASE, has been used in order to successfully analyze the explanatory force of the models accounting for one of the major transitions in evolution, the transition between prokaryotes and eukaryotes. ASE contends that the explanatory force is gained appealing to a symbiogenetic law and the use of a new theoretical role of symbiosis.

The papers as a whole contend that a pluralist picture of scientific explanation in biology is better suited to account for the explanatory practices and models in this scientific discipline. A *weak explanatory pluralism* has been claimed to be the picture that best captures the nature of explanations in biology. I have argued that three accounts of explanation are needed in order to analyze the explanatory practices of biologists: mechanistic explanations, structural explanations and covering law explanations (ASE).

Although the papers of this thesis dissertation help to clarify some issues of the debate about the scope and limits of mechanistic explanation, there is still work to be done. Among the many open questions of the debate, like the validity of *dynamic mechanistic explanations*, or the possibility of integrating topological explanations with mechanistic explanations, I think that the fact of considering covering law accounts and structural accounts of explanation as complementary to mechanistic ones is not yet fully accepted among the philosophers of biology.

On the one hand, the notion of structural explanation was introduced less than ten years ago and is becoming increasingly more used in debates about scientific explanation, although there are still voices among New Mechanists rising doubts about the explanatory force of mathematical modelling. However, the appearance, in the last ten years or so, of a remarkable number of research papers discussing the explanatory role that mathematics plays in biological models shows that there is a current concern amongst philosophers of biology about this topic. Systems biology is a promising new discipline in which the presence of these types of explanation, in particular topological explanation, seems to be pervasive. Thus, I believe analyzing the explanatory practices of systems biology will probably help identify more cases of structural explanations. At the same time, it will contribute to clarify and consolidate the notion of structural explanation as a successful account of explanation in biology. Equilibrium explanations have been comparatively less discussed in the literature on scientific explanation in biology, even though the appeal to equilibrium states is pervasive all-around biology. Again, I think more examples of analysis of equilibrium explanations would help both clarify and consolidate structural explanations in the field.

On the other hand, covering law accounts have had a minimal weight in analyzing explanations in biology in the last few years. However, I contended the notion of pragmatic law of science, a promising notion to account for the nature and role of scientific generalizations, and ASE, a minimal neo-Hempelian account of explanation, might help to recover the importance of covering law accounts of scientific explanation. Unfortunately, the success of the New Mechanism philosophy challenges the possibility that laws are needed in biology, for they articulate the presence of regularities in science in terms of mechanisms. I think that more work on pragmatic laws and ASE is needed to appreciate the important insights that covering law models provide regarding explanation in biology. There is yet an open question, related to this latest point, which has not been addressed in the present dissertation. Topological explanations, and maybe even all types of structural explanation, always display some kind of regularities. For instance, it can be argued that all systems displaying a Small World topology will be robust to perturbations, whatever the nature of the systems is. In light of this, it is reasonable to ask whether ASE captures elements that can be as well present in all instances of topological explanations. An advantage of ASE is that its requirements for there to be scientific explanations are minimal; hence, many types of explanations can fit in its description. If we can identify some kind of regularity and some new conceptual or ontological machinery used to explain in all instances of topological explanations, we could actually argue that topological explanations are not but another version of covering law explanations, thereby aiming for some kind of unification. I will devote future researches to achieve such an aim.

Furthermore, by addressing the questions the papers of this dissertation were asking, I have contributed to clarify notions such as mechanistic explanation, topological explanation, equilibrium explanation and pragmatic laws of science, among others.

Finally, I believe that even if we identify some parts of biology in which mechanistic explanation is not useful to capture the explanatory power of biological models, explanation in many other branches seems to work by integrating mathematical and mechanistic knowledge. This is why, in my opinion, philosophers of biology need to devote more research in order to achieve a whole understanding of how mathematical and mechanistic strategies are integrated in the explanatory practices of scientists.

4. REFERENCES

Alleva, K., Díez, J. A., & Federico, L. (2017). Models, theory structure and mechanisms in biochemistry: The case of allosterism. *Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences*, 63, 1–14.

Andersen, H. (2014). A field guide to mechanisms: Part I. *Philosophy Compass*, *9*(4), 274-283.

Armstrong, D., (1983). *What is a law of nature?* Cambrigde U.P., Cambridge

Bachmaier, C., Brandes, U. & Schreiber, F. (2013). Biological networks. In R. Tamassia (Ed.), Handbook of graph drawing and visualization (pp. 621–651). Hoboken: CRC Press.

Balzer, W., Moulines, C. U., & Sneed, J. D. (1987). *An architectonic for science: The structuralist program*. Reidel. Dordrecht.

Batterman, R. W. (2002). Asymptotics and the role of minimal models. *The British Journal for the Philosophy of Science*, *53*(1), 21-38.

Batterman, R.W. & Rice, C.C. (2014). Minimal Model Explanations. *Philosophy of Science*, 81, 349–376.

Beatty, J. (1995). The evolutionary contingency thesis. In Sober (Ed.), (1995). *Conceptual issues in evolutionary biology*, Mit Press.

Bechtel, W. & Abrahamsen, A. (2005). Explanation: A mechanist alternative. *Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences* 36 (2): 421–441.

Bechtel, W. & Abrahamsen, A. (2010). Dynamic Mechanistic Explanation: Computational Modeling of Circadian Rhythms as an Exemplar for Cognitive Science. *Studies in History and Philosophy of Science* Part A, 41(3), 321–333.

Bechtel, W. & Abrahamsen, A. (2011). Complex Biological Mechanisms: Cyclic, Oscillatory, and Autonomous. In C.A. Hooker (ed.) Philosophy of Complex Systems. Handbook of the Philosophy of Science, Volume 10, New York: Elsevier, 257–285

Bechtel, W. & Abrahamsen, A. (2012). Diagramming phenomena for mechanistic explanation. In *Proceedings of the Annual Meeting of the Cognitive Science Society* (Vol. 34, No. 34).

Bechtel, W., & Abrahamsen, A. A. (2013). Thinking dynamically about biological mechanisms: Networks of coupled oscillators. *Foundations of Science*, *18*(4), 707-723.

Bechtel, W. & Richardson, R.C. (1993). *Discovering complexity: Decomposition and localization as scientific research strategies*. Cambridge, MA: The MIT Press.

Blaser, M. J., & Kirschner, D. (2007). The equilibria that allow bacterial persistence in human hosts. *Nature*, *449*(7164), 843.

Braillard & Malaterre (Eds.), *Explanation in Biology: An Enquiry into the Diversity of Explanatory Patterns in the Life Sciences.* Springer, Dordrecht.

Braillard, P. & Malaterre, C. (2015). Explanation in biology: An introduction. In Braillard & Malaterre (Eds.), *Explanation in Biology: An Enquiry into the Diversity of Explanatory Patterns in the Life Sciences.* pp. 1-28.

Brandon, R. N. (1997). Does biology have laws? The experimental evidence. *Philosophy of Science*, *64*, S444-S457.

Brigandt, I. (2010). Beyond reduction and pluralism: Toward an epistemology of explanatory integration in biology. Erkenntnis, 73(3), 295–311.

Brigandt, I. (2013a). Explanation in biology: Reduction, pluralism, and explanatory aims. *Science and Education*, 22(1), 69–91.

Brigandt, I. (2013b). Systems biology and the integration of mechanistic explanation and mathematical explanation. *Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences*, 44, 477–492.

Brigandt, J. (2015). Evolutionary developmental biology and the limits of philosophical accounts of mechanistic explanation. In P. A. Braillard & C. Malaterre (Eds.), *Explanation in Biology: An Enquiry into the Diversity of Explanatory Patterns in the Life Sciences* (pp. 135–173). Dordretcht: Springer.

Brigandt, J., Green, S., & O'Malley, M. A. (2018). Systems biology and mechanistic explanation. In Glennan, S. & Illari, P. (Eds.). (2017). *The Routledge handbook of mechanisms and mechanical philosophy*. Taylor & Francis. Chapter 27.

Carnap, R. (1966). *Philosophical foundations of physics* (Vol. 966). New York: Basic Books.

Carrier, M. (1995). Evolutionary change and lawlikeness: Beatty on biological generalizations. *Concepts, theories and rationality in the biological sciences*, 82-97.

Carroll, J. W. (2016). Laws of Nature. *The Stanford Encyclopedia of Philosophy* , Edward N. Zalta (ed.), https://plato.stanford.edu/archives/fall2016/entries/laws-of-nature/

Cartwright, N. (1979). Causal laws and effective strategies. *Noûs,* 13, 419–437.

Cartwright, N. (1980). Do the laws of physics state the facts? *Pacific Philosophical Quarterly*, 61, 75–84.

Chemero, A. & Silberstein, M. (2008). After the philosophy of mind: Replacing scholasticism with science. *Philosophy of science* 75.1: 1-27.

Cohen, J. & Callender, C. (2009). A better best system account of lawhood. *Philosophical. Studies* 145.1: 1-34.

Craver, C. F. (2006). When mechanistic models explain. *Synthese*, *153*(3), 355-376.

Craver, C. F. (2007). *Explaining the brain*. New York: Clarendon Press.

Craver, C. F. & Darden, L. (2005). Introduction: Mechanisms Then and Now. *Special Issue:"Mechanisms in Biology," Studies in History and Philosophy of Biological and Biomedical Sciences*, *36*, 233-244.

Craver, C. F. & Darden, L. (2013). *In search of mechanisms: Discoveries across the life sciences*. University of Chicago Press.

Craver, C. F. & Kaiser, M. I. (2013). Mechanisms and laws: Clarifying the debate. In *Mechanism and causality in biology and economics* (pp. 125-145). Springer, Dordrecht.

Craver, C. F. & Tabery, J. (2015). Mechanisms in Science in E.N. Zalta (ed.). *The Stanford Encyclopedia of Philosophy* https://plato.stanford.edu/entries/science-mechanisms/

Darden, L. (2008). Thinking again about biological mechanisms. *Philosophy of science*, *75*(5), 958-969.

Darden, L. & Craver, C. F. (2002). Strategies in the interfield discovery of the mechanism of protein synthesis. *Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences* 33.1. pp.1-28.

Deulofeu, R., & Suárez, J. (2018). When mechanisms are not enough: the origin of eukaryotes and scientific explanation. In Christian, A., Hommen, D., Retzlaff, N. & Schurz, G. (editors). (2018). *Philosophy of Science. Between the natural sciences, the social sciences and the humanities.* European Studies in Philosophy of Science. Volume 9.

Deulofeu, R., Suárez, J. & Pérez-Cervera, A. (2019). Explaining the behavior of random ecological networks: the stability of the microbiome as a case of integrative pluralism. *Synthese,* 1-23.

Diéguez, A. (2012). *La vida bajo escrutinio: una introducción a la Filosofía de la Biología*. Ediciones de Intervención Cultural.

Díez, J. A. (2014). Scientific w-Explanation as Ampliative, Specialized Embedding: a neo-Hempelian account. *Erkenntnis*, *79*(8), 1413-1443.

Díez, J. A., Lorenzano, P. (Eds.), (2002). *Desarrollos actuales de la metateoría estructuralista: problemas y discusiones*. Universidad Nacional de Quilmes Ediciones.

Díez, J. A. & Lorenzano, P. (2013). Who got what wrong? Fodor and Piattelli on Darwin: guiding principles and explanatory models in natural selection. *Erkenntnis*, *78*(5), 1143-1175.

Díez, J.A. & Lorenzano, P. (2015). Are natural selection explanatory models a priori?. *Biology & Philosophy*, *30*(6), 787-809.

Díez, J. A. & Moulines, C. U. (2008). *Fundamentos de filosofía de la ciencia* (No. 165). Ariel

Dorato, M. (2012). Mathematical biology and the existence of biological laws. In *Probabilities, Laws, and Structures* (pp. 109-121). Springer, Dordrecht.

Dowe, P. (2000). Physical Causation, Cambridge: Cambridge University Press.

Dretske, F. (1977). Laws of nature. *Philosophy of science*: 248-268

Garson, J. (2018). Mechanisms, phenomena, and functions. In Glennan, S. & Illari, P. (Eds.). (2017). *The Routledge handbook of mechanisms and mechanical philosophy*. Taylor & Francis. Chapter 8.

Glennan, S. (1996). Mechanisms and the Nature of Causation. *Erkenntnis* 44:50–71.

Glennan, S. (2002). Rethinking mechanistic explanation. *Philosophy of Science* 69 (S3): S342–S353.

Glennan, S. (2017). *The new mechanical philosophy*. Oxford University Press.

Glennan, S. & Illari, P. (Eds.). (2018). *The Routledge handbook of mechanisms and mechanical philosophy*. Taylor & Francis.

Gould, S. J. (1989). *Wonderful life: the Burgess Shale and the nature of history*. WW Norton & Company.

Green, S. (2017). *Philosophy of Systems Biology*. Dordrecht: Springer.

Green, S. (2019). Philosophy of Systems and Synthetic Biology. *The Stanford Encyclopedia of Philosophy* (Spring 2019 Edition), Edward N. Zalta (ed.), https://plato.stanford.edu/archives/spr2019/entries/systems-synthetic-biology/.

Halina, M. (2018). Mechanistic explanations and its limits. In Glennan, S. & Illari, P. (Eds.). (2017). *The Routledge handbook of mechanisms and mechanical philosophy*. Taylor & Francis. Chapter 16

Hempel, C. G. (1962). Deductive-nomological vs. statistical explanation. In Feigl & Maxwell (eds.) (1962), *Minnesota Studies in the Philosophy of Science*. III, U. Minnesota P., Minneapolis, 98-170.

Hempel, C.G. (1965). *Aspects of scientific explanation and Other Essays in the Philosophy of Science.* Free P., New York.

Hempel, C. G. & Oppenheim, P. (1948). Studies in the Logic of Explanation. *Philosophy of science*, *15*(2), 135-175.

Huneman, P. (2010). Topological explanations and robustness in biological sciences. *Synthese,* 177, 213–245.

Huneman, P. (2018a). Outlines of a theory of structural explanation. *Philosophical Studies*, 175(3), 665–702.

Huneman, P. (2018b). Diversifying the picture of explanations in biological sciences: Ways of combining topology with mechanisms. *Synthese*, 195, 115–146.

Huneman, P. (2018c). Realizability and the varieties of explanation. *Studies in History and Philosophy of Science.* https://doi.org/10.1016/j.shpsa.2018.01.004.

Illari, P. M. & Williamson, J. (2012). What is a mechanism? Thinking about mechanisms across the sciences. *European Journal for Philosophy of Science*, *2*(1), 119-135.

Issad, T. & Malaterre, C. (2015). Are Dynamic Mechanistic Explanations Still Mechanistic? *Explanation in Biology: An Enquiry into the Diversity of Explanatory Patterns in the Life Sciences*. Chapter 12. Springer, Dordrecht.

Jones, N. (2014). Bowtie structures, pathway diagrams, and topological explanations. *Erkenntnis,* 79(5),
1135–1155.

Kaiser, M.I. (2018). The components and boundaries of mechanisms. In Glennan, S. & Illari, P. (Eds.). (2017). *The Routledge handbook of mechanisms and mechanical philosophy*. Taylor & Francis. Chapter 9.

Kaplan, M.D. (2018). Mechanisms and dynamical systems. In Glennan, S. & Illari, P. (Eds.). (2017). *The Routledge handbook of mechanisms and mechanical philosophy*. Taylor & Francis. Chapter 20.

Kaplan, D. M. & Craver, C.F. (2011). The explanatory force of dynamical and mathematical models in neuroscience: A mechanistic perspective. *Philosophy of science* 78.4 (2011): 601-627.

Kaplan, D.M. & Bechtel, W. (2011). Dynamical models: An alternative or complement to mechanistic explanations?. *Topics in Cognitive Science* 3.2: pp. 438-444.

Kellert, S.H., Longino, H.E. and Waters, C.K. (2006). Introduction: The Pluralist Stance. in S.H. Kellert, H.E. Longino, and C.K. Waters (eds.) *Scientific Pluralism*, Minneapolis: University of Minnesota Press, vii–xxix.

Kitano, H. & Oda, K. (2006). Robustness trade-offs and host–microbial symbiosis in the immune system. *Molecular systems biology*, *2*(1).

Kitcher, P. (1989). Explanatory unification and the causal structure of the world. In P. Kitcher & W. Salmon (Eds.), Scientific explanation (pp. 410–505). Minneapolis: University of Minnesota Press.

Kitcher, P. (1993). The Advancement of Science. Oxford: Oxford University Press.

Kitcher, P. (1999). Unification as a regulative ideal. *Perspectives on Science*, 7, 337–348.
Krickel, B. (2014), The Metaphysics of Mechanism, PhD Dissertation. Humboldt-Universität zu Berlin.

Kuorikoski, J. (2007). Explaining with equilibria. In *Rethinking explanation* (pp. 149-162). Springer, Dordrecht.

Lange, M. (2013). Really statistical explanations and genetic drift. *Philosophy of Science*, *80*(2), 169-188.

Levy, A. (2013). Three kinds of new mechanism. *Biology & Philosophy*, *28*(1), 99-114.

Leuridan, B. (2010). Can mechanisms really replace laws of nature?. *Philosophy of Science*, *77*(3), 317-340.

Lorenzano, P. (2007). Leyes fundamentales y leyes de la biología. *Scientiae Studia* 5 (2):185-214.

Love, A. C. (2008). Explaining evolutionary innovations and novelties: Criteria of explanatory adequacy and epistemological prerequisites. *Philosophy of Science*, 75, 874–886.

Machamer, P., Darden, L. & Craver, C.F. (2000). Thinking about mechanisms. *Philosophy of Science* 67 (1): 1–25.

Mekios, C. (2015). Explanation in systems biology: is it all about mechanisms? *Explanation in Biology: An Enquiry into the Diversity of Explanatory Patterns in the Life Science.* Chapter 3. Springer, Dordrecht.

Militello, G., & Moreno, Á. (2018). Structural and organisational conditions for being a machine. *Biology & Philosophy*, *33*(5-6), 35.

Mitchell, S. D. (1997). Pragmatic laws. *Philosophy of Science*, *64*, S468-S479.

Mitchell, S.D. (2000). Dimensions of scientific law. *Philosophy of Science* 67 (2): 242–265.

Mitchell, S. D. (2002). Integrative pluralism. *Biology and Philosophy*, *17*(1), 55-70.

Mitchell, S.D. (2003). *Biological complexity and integrative pluralism*. Cambridge: Cambridge University Press

Montoya, J. M. & Solé, R. V. (2002). Small world patterns in food webs. *Journal of theoretical biology*, *214*(3), 405-412.

Moreno A., and J. Suárez (submitted). "Plurality of Explanatory Strategies in Biology: Mechanisms and Networks".

Nagel, E. (1961). *The structure of science*. Harcourt. New York

Nicholson, D.J. (2012). The concept of mechanism in biology. *Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences* 43 (1): 152–163.

Nikisianis, N. & Stamou, G. P. (2016). Harmony as ideology: Questioning the diversity-stability hypothesis. *Acta Biotheoretica*, 64(1), 33–64.

O'Malley, M. A. & Soyer, O. S. (2012). The roles of integration in molecular systems biology. *Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences*, *43*(1), 58-68.

Potochnik, A. (2015). Causal patterns and adequate explanations. *Philosophical Studies*, *172*(5), 1163-1182.

Rice, C. (2015). Moving beyond causes: Optimality models and scientific explanation. *Noûs*, *49*(3), 589-615.

Ross, L. (2015). Dynamical Models and Explanation in Neuroscience. *Philosophy of Science*, 82(1), 32–54.

Rosenberg, A. (2001). How is biological explanation possible?.*The British Journal for the Philosophy of Science*, *52*(4), 735-760.

Salmon, W. C. (ed.). (1971). *Statistical Explanation and Statistical Relevance*. Pittsburgh: University of Pittsburgh Press.

Salmon, W. C. (1984). *Scientific explanation and the causal structure of the world*. Princeton: Princeton University Press.

Salmon, W. C. (1989). Four decades of scientific explanation. *Scientific explanation*, *13*, 3-219.

Salmon, W. C. (1994). Causality without counterfactuals. *Philosophy of Science*, 61, 297–312.

Strevens, M. (2004). The causal and unification approaches to explanation unified. *Noûs*, *38*(1), 154-176.

Scriven, M. (1962). Explanations, predictions, and laws. In H. Feigl & G. Maxwell (Eds.), Minnesota studies in the philosophy of science, scientific explanation, space, and time (Vol. 3, pp. 170–230). Minneapolis: University of Minnesota Press.

Skipper Jr, R. A. & Millstein, R.L. (2005). Thinking about evolutionary mechanisms: Natural selection. *Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences* 36.2 (2005): 327-347.

Smart, J. J. C. (2014). *Philosophy and scientific realism*. Routledge.

Smith, J. M. & Price, G. R. (1973). The logic of animal conflict. *Nature*, *246*(5427), 15.

Sober, E. (1983). Equilibrium explanation. *Philosophical Studies*, *43*(2), 201-210.

Solé, R. V. & Montoya, J. M. (2001). Complexity and fragility in ecological networks. *Proceedings of the Royal Society B,* 268(1480), 2039–2045.

Stepp, N., Chemero, A. & Turvey, M. T. (2011). Philosophy for the rest of cognitive science. *Topics in Cognitive Science* 3.2: 425-437.

Strogatz, S. H. (2001). Exploring complex networks. *Nature* 410, 268-275

Suárez, J. & Deulofeu, R. (Forthcoming). Equilibrium explanations as structural non-mechanistic explanations: the case of long-term bacterial persistence in human hosts. *Teorema.*

Théry, F. (2015). Explaining in contemporary molecular biology: Beyond mechanisms. *Explanation in Biology: An Enquiry into the Diversity of Explanatory Patterns in the Life Science.* Chapter 6. Springer, Dordrecht.

Van Fraassen, Bas C. (1980). *The scientific image*. Oxford University Press.

Walsh, D. M. (2015). Variance, Invariance, and Statistical Explanation. *Erkenntnis* 80: 469–489.

Walsh, D. M., Lewens, T. & Ariew, A. (2002). The Trials of Life: Natural Selection and Random Drift. *Philosophy of Science* 69 (3): 429–46.

Walsh, D. M., Ariew, A & Matthen, M. (2017). Four pillars of statisticalism. *Philosophy, Theory & Practice in Biology.* Volume 9. Nº1.

Watts, D. J. & Strogatz, S. H. (1998). Collective dynamics of 'small-world' networks. *Nature*, *393*(6684), 440.

Woodward, J. (1989). The Causal/Mechanical Model of Explanation.', In *Scientific Explanation.* Minnesota Studies in the Philosophy of Science, volume 13, P. Kitcher and W. Salmon (eds.), 357–383.

Woodward, J. (2002). What is a mechanism? A counterfactual account. *Philosophy of Science*, *69*(S3), S366-S377.

Woodward, J. (2003). *Making things happen*. New York: Oxford University Press.

Woodward, J. (2013). II—Mechanistic explanation: Its scope and limits. In *Aristotelian Society Supplementary Volume* (Vol. 87, No. 1, pp. 39-65). Oxford, UK: Blackwell Publishing Ltd.

Woodward, J. (2016). Causation and Manipulability. *The Stanford Encyclopedia of Philosophy* (Winter 2016 Edition), Edward N. Zalta (ed.), https://plato.stanford.edu/archives/win2016/entries/causation-mani/

Woodward, J. (2017). Scientific Explanation. *The Stanford Encyclopedia of Philosophy* (Fall 2017 Edition), Edward N. Zalta (ed.), https://plato.stanford.edu/archives/fall2017/entries/scientific-explanation/

Xia, T. (2010). Network modelling in systems biology. Ph.D. Dissertation. https://lib.dr.iastate.edu/cgi/viewcontent.cgi?article=2611&context=etd.

II. ANNEX

- Deulofeu, R., & Suárez, J. (2018). When mechanisms are not enough: the origin of eukaryotes and scientific explanation. In Christian, A., Hommen, D., Retzlaff, N. & Schurz, G. (editors). (2018). *Philosophy of Science. Between the natural sciences, the social sciences and the humanities.* European Studies in Philosophy of Science. Volume 9.

- Deulofeu, R., Suárez, J. & Pérez-Cervera, A. (2019). Explaining the behavior of random ecological networks: the stability of the microbiome as a case of integrative pluralism. *Synthese,* 1-23.

- Suárez, J. & Deulofeu, R. (2019). Equilibrium explanations as structural non-mechanistic explanations: the case of long-term bacterial persistence in human hosts. *Teorem: revista internacional de filosofía,* 38 (3), 95-120

Chapter 6
When Mechanisms Are Not Enough: The Origin of Eukaryotes and Scientific Explanation

Roger Deulofeu and Javier Suárez

Abstract The appeal to mechanisms in scientific explanation is commonplace in contemporary philosophy of science. In short, mechanists argue that an explanation of a phenomenon consists of citing the mechanism that brings the phenomenon about. In this paper, we present an argument that challenges the universality of mechanistic explanation: in explanations of the contemporary features of the eukaryotic cell, biologists appeal to its symbiogenetic origin and therefore the notion of symbiogenesis plays the main explanatory role. We defend the notion that symbiogenesis is non-mechanistic in nature and that any attempt to explain some of the contemporary features of the eukaryotic cell mechanistically turns out to be at least insufficient and sometimes fails to address the question that is asked. Finally, we suggest that symbiogenesis is better understood as a pragmatic scientific law and present an alternative non-mechanistic model of scientific explanation. In the model we present, the use of scientific laws is supposed to be a minimal requirement of all scientific explanations, since the purpose of a scientific explanation is to make phenomena expectable. Therefore, this model would help to understand biologists' appeal to the notion of symbiosis and thus is shown to be better, for the case under examination, than the mechanistic alternative.

Keywords Scientific explanation · Mechanistic explanation · Scientific laws · Eukaryotic cell · Symbiogenesis · Symbiosis

This work is fully collaborative; the authors are listed alphabetically.

R. Deulofeu (✉)
Department of Philosophy, University of Barcelona, Barcelona, Spain
e-mail: rogerdeulofeu@ub.edu

J. Suárez
Department of Sociology, Philosophy and Anthropology, University of Exeter, Exeter, UK
e-mail: javier.suarez@exeter.ac.edu

A. Christian et al. (eds.), *Philosophy of Science*, European Studies in Philosophy of Science 9, https://doi.org/10.1007/978-3-319-72577-2_6

6.1 Introduction

In recent years, mechanistic talk has become very popular among philosophers of science. Particularly, mechanistic talk has displaced the traditional approach to scientific explanation in terms of scientific laws (Nicholson 2012). Mechanists claim that scientific explanation consists of looking for a *causal process* –in this sense, the mechanistic movement is just the other side of the coin of traditional causal models of explanation– such that, through connecting the different entities and activities that participate in the process, the phenomenon that we aim to explain simply emerges. This claim is in contrast with the claim made by defenders of nomological expectability models of scientific explanation who generally claim that "to explain a phenomenon is to make it expectable on the basis of non-accidental regularities" (Díez 2014, 1414). Mechanists usually put forward biology as their main counterexample against defenders of nomological models: when biologists claim to have explained a phenomenon, they do so on the basis of having found a mechanism that brings that phenomenon about (Machamer et al. 2000). Biologists do not appeal to laws of nature, logical arguments, or any other kind of logic: they simply appeal to mechanisms. Thus, scientific explanation is, on this view, mechanistic explanation. In this paper, we contend this claim on its own terms, by presenting an example from biological practice. Specifically, we present the case of the origin of the eukaryotic cell and argue that the explanation of the salient features of this peculiar case is more suited to be understood in terms of a nomological expectability model of scientific explanation than in terms of mechanisms. For this purpose, we make explicit a kind of general regularity that biologists seem to be assuming when they provide explanations of the origin of the eukaryotic cell, and which forms the basis of the kind of proposals that they take as explanatory of certain facts that they consider particularly salient and in need of explanation (see Alleva et al. 2017, for a similar line of reasoning applied to the case of allosterism).

The paper is organised as follows: In Sect. 6.2, we introduce the symbiosis theory (ST, hereafter) of the origin of the eukaryotic cell, nowadays considered the canonical model for explaining the origin of eukaryotic cells, and we introduce a classification of the questions that ST provides answer to. In Sect. 6.3, we introduce the mechanistic account of scientific explanation complemented with Woodward's account of causality and provide evidence that suggests that the appeal to mechanisms is not the most appropriate way to justify the explanatory character of ST of the origin of the eukaryotic cell and why this is so. In Sect. 6.4, we present a nomological expectability model of scientific explanation that we then use to provide an understanding of the explanatory character of the ST of the origin of eukaryotic cells by considering that ST appeals to scientific laws. Finally, in Sect. 6.5 we conclude by defending the superiority of the nomological approach over the mechanistic approach in providing an understanding of the explanatory practices of biologists in the context of the theories of the origin of the eukaryotic cell and we propose future lines of research.

6.2 Symbiosis Theories of the Origin of Eukaryotic Cells

The biological world is populated by different kinds of entities, ranging from cells, to all kinds of multicellular forms of life. Cells are normally taken to be the basic and most fundamental unit of life, of which all the other entities are made up (Archibald 2014; Audesirk et al. 2008; Stearns and Hoekstra 2000). There are two main types of cells, classified according to the location of their DNA: prokaryotic cells (subdivided into the domains of Archaea and Bacteria) and eukaryotic cells. The main structural difference between prokaryotic cells and eukaryotic cells is that in the former, the genetic material is dispersed throughout the cytoplasm; whereas in the latter it is encapsulated within a membranoid-structure called the "nucleus". Apart from this, there are many other structural differences between the two types of cells, concerning aspects such as their size (eukaryotic cells generally being bigger), the types of membranes and the presence or absence of organelles. This last different constitutes a salient feature of eukaryotic cells, since only they host organelles within their bodies. Organelles are structural subunits, analogous to organs in humans, which perform certain functions within the body of the cell they belong to. Two of the organelles within eukaryotic cells are mitochondria (present in all eukaryotic cells) and chloroplasts (present only in plant eukaryotic cells); these two organelles bear their own DNA. Mitochondria are the site of cell respiration. Photosynthesis, in contrast, takes places within chloroplasts. Eukaryotic and prokaryotic cells are quite distinct from each other, and there does not seem to be any record of an intermediate form between the two types of cells, which is why certain biologists have referred to the origin of the eukaryotic cells as "the greatest single evolutionary discontinuity to be found in the present-day living world" (Stainer et al. 1963, quoted in Sagan 1967, 226). This immediately triggers a serious question for biologists: how did the first eukaryotic cell appear, given that all organisms share a common ancestor, and therefore eukaryotes and prokaryotes must have originated from the same ancestor?

Answering this question about the origin of the eukaryotic cell consists, among other things, of explaining the origin of cellular organelles, as the most salient subunits that allow for the distinction between eukaryotes and prokaryotes, and particularly of answering questions about the origin of mitochondria and chloroplasts. Mitochondria and chloroplasts are, then, one of the hallmarks of "eukaryocity" and, as Martin and his collaborators have put it, "the invention of eukaryotic specific traits required more metabolic energy per gene than prokaryotes have at their disposal, and (...) mitochondria afforded eukaryotic cells an order of magnitude increase in the amount of energy per gene, which (finally) explains *why* the origin of eukaryotes corresponds to the origin of mitochondria" (Martin et al. 2015, 2; also Williams and Embley 2015, Sect. 6.1).[1] Furthermore, it consists of justifying

[1]This point is however controversial, as some people have also defended the idea that other processes such as a phagocytosis might also be considered as the starting point of eukaryocity (e.g. Cavalier-Smith 1989). However, that would not remove the need to explain the origin of

the lack of continuity in the fossil record between eukaryotes and prokaryotes, the biochemical differences between the two types of cells, the different capabilities of one type of cells with respect to the other, etc. Explaining the origin of eukaryotic cells consists, therefore, of providing satisfactory answers to a series of why-questions (facts) about the particular features of the two kinds of cells and especially answering certain questions about the particular nature of each type. The family of surprising facts that a theory of the origin of the eukaryotic cell has to provide explanations of can be roughly classified as:

- *Physiological and biochemical questions*. The model of the origin of the eukaryotic cells has to explain, for instance, why the membrane of mitochondria is biochemically quite distinct from the membrane of the eukaryotic cell, but biochemically closely related to the nature of the membranes of certain prokaryotes; it also has to explain why the genetic material of eukaryotes has a mosaic nature, i.e. it is composed of phylogenetically distinct classes of DNA.
- *Phylogenetic questions*. Mitochondria and chloroplasts are not phylogenetically close to eukaryotes, but they are phylogenetically close to certain prokaryotes. This fact is surprising, since mitochondria are organelle in the eukaryotic cell, so one important question to answer would be why their genetic material is distinct in nature from the one present in the eukaryotic nucleus.
- *Historical questions*. The most important question to be answered is why there is a gap in the fossil record between prokaryotes and eukaryotes, if we take evolution to be continuous with no sudden evolutionary jumps.

So, a theory of the origin of the eukaryotes (i.e. a theory that answers the question: "How did eukaryotic cells originate?") should provide satisfactory answers to a list of why-questions of different natures, and evaluating its success at doing so is fundamental for the acceptance of one theory over another.

To answer the set of question outlined above, two families of theories have been proposed: on the one hand, *self-genetic or autogenous theories*, according to which the organelles within eukaryotes appeared as a consequence of invaginations within the original pre-eukaryotic cell (Raff and Mahler 1972; Uzzel and Spolsky 1974; all reviewed in Sapp 2010, 130–131; O'Malley 2010; Archibald 2015, R912); and on the other, *symbiosis or exogenous theories*, whose main claim is that eukaryotic cells originated through the symbiotic merger of two previously extant prokaryotic cells (Margulis 1970; Martin et al. 2012; Cavalier-Smith 2013; Dolan 2013). In short, the proponents of ST argue that the eukaryotic cell evolved as a consequence of a phagocytic process in which prokaryotes "were swallowed but not digested" (Margulis 1970, 60). The difference between the two families of theories is radical, and so are the conclusions that one can derive from them. For instance, if one

mitochondria and chloroplasts in a satisfactory manner, which would lead to the same kind of questions that we mention later. For the purposes of this paper and for simplicity, we will follow Martin's proposal that equates the origin of eukaryotes with the origin of mitochondria. Thanks to Thomas Bonnin for pointing this out to us.

defends an autogenous theory, one has difficulties explaining the genetic affinities between mitochondrial DNA and the DNA of free-living prokaryotes, since one has to explain how this foreign DNA arrived in the mitochondria of present-day eukaryotes. However, if one defends a ST, this fact becomes easily explainable: the fact that in the origin of eukaryotes two different prokaryotic lineages merged makes it more likely that the primitive lineage associated with mitochondria still preserves part of its original DNA. The same logic can be applied to all kinds of questions that might be raised about the difference between prokaryotes and eukaryotes. So, the capacity to play a more effective role in scientific explanation proves to be a good feature for preferring one theory to another.

Nowadays, the ST family predominates among biologists, although the versions of it come in many different forms, with at least 20 different models that explain the origin of the eukaryotic cells appealing to symbiosis (Archibald 2015). What matters for the purposes of this paper is the general structure of the arguments that appeal to endosymbiosis to explain the origin of eukaryotes and to explain the set of why-questions that we have selected as relevant, more than the peculiarities of the different models.

In general, ST appeals to the notion of *symbiogenesis* as the process by which the eukaryotic cell originally appeared.[2] This symbiogenetic process is supposed to have given rise to an endosymbiotic relationship between the different interacting organisms. The initial organisms involved in the origin of the first eukaryote are hypothesized to have been an archaeon (although there is no definite consensus on this question), with the capacity to phagocytize other microorganisms, and an alpha-proteobacteria, which would have given rise to mitochondria as we know it today (Spang et al. 2015). The peculiar nature of *symbiogenesis* qualifies it as the reason that biologists offer to explain the surprising features that are observed in eukaryotic cells. For instance:

- Why is the membrane of mitochondria biochemically more similar to free-living proteobacteria than to its host, i.e. the eukaryotic cell itself?
  Because it originated through symbiogenesis, which means that a free-living microorganism was engulfed but not digested and therefore it is very likely that the lineage this previously free-living microorganism gave rise to still preserves some of its original biochemical properties, such as the composition of the membrane.

[2]Symbiogenesis is the process of generation of a new biological structure (organ, metabolic pathway, etc.) as a consequence of a long-term symbiotic association. In the case of the eukaryotic cell, symbiogenesis refers to the origin of a complete new biological domain as a consequence of symbiosis. Symbiotic organisms can interact in two different ways: endosymbiotically, if one organism lives within the cell(s) of the other, and ectosymbiotically, when one organism lives on the surface of the cell(s) of the other, but not within them (Archibald 2014). Symbiogenesis is thus a process, whereas endosymbiosis is a state. This distinction has to be kept in mind for the rest of the paper. Thanks to an anonymous reviewer for encouraging us to clarify this point.

- Why does the eukaryotic genome have a mosaic nature?
  Because it originated by symbiogenesis. This entails two free-living organisms suddenly stopping their free-living mode to live together as a unit. As a consequence of a long-term relationship after symbiogenesis, it is very likely that there will be genetic exchange between the partners, thereby creating the mosaic structure of the eukaryotic genome.
- Why are mitochondria phylogenetically closer to free-living alpha-proteobacteria than to their host?
  Because if mitochondria were once free-living microorganisms that, via a process of symbiogenesis, became organelles within the eukaryotic cell, it seems natural that their DNA would be phylogenetically closer to the DNA of the free-living forms from which they originated than to eukaryotic DNA.
- Why is there a gap in the fossil record between prokaryotes and eukaryotes?
  Because if eukaryotic cells appeared through symbiogenesis, it is very unlikely that intermediate forms would be found in the fossil record. Symbiogenesis is a discontinuous process.

The appeal to symbiogenesis is therefore used as a general strategy to answer a different set of why-questions concerning particular features of the eukaryotic cell, providing answers that trace these features back to their evolutionary origin. In the following sections, we analyse whether this general strategy used by biologists is more in accordance with a mechanistic theory of explanation or with a nomological expectability approach, and we argue that what biologists actually do seems closer to the latter.

6.3 Mechanistic Explanation

Mechanistic explanation is the most influential approach to explanation in biology. The view was originally presented in direct opposition to the (previously) dominant nomological models of scientific explanation. Mechanists argue that in order to explain a biological phenomenon it is necessary to describe the mechanism that brings the phenomenon about (Glennan 1996, 2002; Bechtel 2011; Bechtel and Richardson 1993; Bechtel and Abrahamsen 2005; Machamer et al. 2000; Craver 2006, 2007; Darden and Craver 2002). Describing a mechanism, they claim, is not the same as presenting a scientific law that underlies a phenomenon. In fact, they deny the possibility of explaining a phenomenon by subsuming it under laws. In other words, the explanatory character of a mechanism does not lie on its supposedly underlying regularities, but in the identification of causal relations: “while it is sometimes the case that description of the inner parts of the mechanism will entail a description of the mechanism’s outward behaviour, the explanation lies not in the logical relation between these descriptions but in the causal relations between the

parts of the mechanism that produce the behaviour described" (Glennan 2002, S348; see also Machamer et al. 2000 for a similar argument).[3]

There are several ways of describing what a mechanism is. For instance, Machamer et al. (2000, 3, our emphasis) claim that a mechanism is a set of "*entities* and *activities organized* such that they are productive of regular changes from starting or set-up conditions to finish or termination conditions"; Glennan (2002, S344, our emphasis) defines a mechanism by saying that it is a "complex system that produces the behavior by the *interaction* of a number of *parts*"; Bechtel (2006, 26, our emphasis) says that it is "a structure performing a function in virtue of its component *parts*, component *operations*, and their *organizations*".

It seems clear from the above definitions that all of them presuppose that a mechanism consists of a set of entities and activities (or parts and operations/interactions) plus their corresponding organization.[4] To identify a mechanism, therefore, one has to disentangle its parts (the entities), individuated by their properties, and the activities it is involved in, "the producers of change". Allegedly, the properties of the entities plus their organization are responsible for the way in which the activities come about. In the words of Machamer et al.: "Mechanisms are identified and individuated by the activities and entities that constitute them, by their start and finish conditions and by their functional roles" (2000, 6). This dualist reading of mechanisms in terms of entities and activities generates a new framework that should, in principle, be fruitful when it comes to clarifying notions such as causation, lawhood, function and explanation. In particular, the notion of activity is supposed to play the role of causes, laws and functions. For instance, if a law is supposed to be a regularity of something that acts in the same way under the same conditions, philosophers of a mechanistic bent can provide a similar reading of a mechanism: "a mechanism is the series of activities of entities that bring about the finish or termination conditions in a regular way" (Machamer et al. 2000, 7). According to such authors, these regular mechanisms are not accidental and can give support to counterfactual reasoning. Therefore, there is no need to talk of laws in biology, for their role is already played by the identification of activities within mechanisms. In the same vein, Glennan refers to the interactions within a mechanism as "*invariant change-relating generalizations*" which can support counterfactual claims (Glennan 2002, S344).

[3]Leuridan (2010) argues that for every mechanism we can find an underlying regularity. His conclusion is that the explanatory character of mechanisms lies precisely in these hidden regularities, which actually is conceding too much to the nomological expectability models mechanists were criticizing in the first place.

[4]In fact, the notions of entities and activities come from a modification of the previous description of a mechanism in terms of parts and operations/interactions. Bechtel and Glennan still define mechanisms by appealing to the notions of parts and operations/interactions. The motives for their choice can be found in Bechtel and Abrahamsen (2005, fn. 5). Machamer et al. (2000, §3) introduced the new notions of entities and activities, mainly for ontological reasons. We take this not to be a substantive distinction for the purposes of this paper.

Given this characterization of a mechanism, we can now say that to give a mechanistic explanation of a given phenomenon consists of giving a description of the mechanism that brings the phenomenon about, such that the explanans includes the set-up conditions (arbitrarily taken as the beginning of the mechanism) plus the intermediate entities and activities together with their organization.

Nonetheless, there still remains the problem of providing criteria for identifying the different parts that compose a mechanism and that should be taken as relevant for the purposes of explanation. One possible way out of this problem, adopted among others by Craver (2007, 144), is to make use of Woodward's manipulability criteria for identifying causes (Woodward 1997, 2000, 2003). Woodward's strategy is to look for a "difference-making" clause in the explanans that, if we were to change it in various possible ways, would result in the final phenomenon being different. This strategy is mainly interventionist: if we want to identify the relevant factors for the production of a particular phenomenon, we must block certain parts allegedly involved in the causal path that terminates in the phenomenon to see whether this intervention has any consequence on the final output. Following this line of reasoning, one can say that "a part is causally relevant to the phenomenon produced by a causal mechanism if one can modify the production of this phenomenon by manipulating the behavior of the part, and one can modify the behavior of the part by manipulating the production of the phenomenon by the causal mechanism" (Nicholson 2012, 160).

Woodward is conscious that the interventions he requires to uncover the causes of phenomena might not always be available (think, for example, of historical phenomena). In order to resolve this difficulty, he argues that in those contexts where such manipulation is not feasible, the manipulability strategy takes the form of a counterfactual claim: "The notion of information that is relevant to manipulation thus needs to be understood modally or counterfactually: the information that is relevant to causally explaining an outcome involves the identification of factors and relationships such that if (perhaps contrary to fact) manipulation of these factors were possible, this would be a way of manipulating or altering the phenomenon in question" (Woodward 2003, 10). In other words, even in contexts where manipulation is not possible, it is "heuristically useful" to pursue or think of causes in the same way as we do when the relevant manipulation is available.

The task now is to try to apply the mechanistic schema plus Woodward's account of causes to the explanation of the origin of the eukaryotic cell in order to test its usefulness. We will advance question by question, following the schema presented in Sect. 6.1:

- Why is the membrane of mitochondria biochemically more similar to free-living proteobacteria than to its host, i.e. the eukaryotic cell itself?
- This question is about similarity, i.e. it is about why certain biomarkers are similar in an organelle and an organism that is phylogenetically distant from the eukaryotic cell that bears the organelle, whereas those biomarkers are different between the eukaryotic cell and its organelle. The mechanist would want to look for the different entities and activities, and their organization that would

allow the phenomenon under investigation (the nature of the membrane) to occur. If we were to do that, the entities would be the membranes and their biochemical nature; the activities would be those of membrane synthesis and membrane destruction; and the organization would depend on the way in which the aforementioned parts are spatiotemporally located in standard mitochondria. Let us suppose we follow this strategy. It is highly likely that we will discover many details about membrane synthesis, the biochemical products that are produced, the way in which they relate to each other, how they become arranged within mitochondria to give rise to a new membrane, etc. However, valuable as this information might be, it does not provide us with the answer we are looking for. This line of research would isolate the causes, allow interventions and provide a better understanding of membrane composition and membrane synthesis. But this is not what we were looking for in the first place. Our question concerned the similarities between mitochondria and a free-living microorganism, and the best answer to the question lies in symbiogenesis, as we mentioned in Sect. 6.1, and nothing in the strategy that the mechanist might elaborate mentions symbiogenesis.

Nevertheless, the mechanicist might still try to argue that the explanation lies in symbiogenesis because symbiogenesis is, in this particular circumstance, a mechanism. The problem is that we are looking for a historical explanation and thus we can only apply Woodward's counterfactual strategy. But this does not seem to do the trick either. First, the notion of symbiogenesis does not look like a mechanism at all: it is a very formal and general notion which does not make any reference to entities (it is supposed to cover a wide range of them, from the eukaryotic cell to most insect microbiota), activities (also very wide and diverse, from oxidation of glucose to synthesis of essential amino acids) or organization (which can be very variable). Second, because of the complexity of symbiogenesis, one cannot even imagine a set of factors whose alteration would block the phenomenon from appearing. If the factor we blocked was the symbiotic merger itself, then the result is not that we do not have a biochemical similarity between mitochondria and certain free-living bacteria: the result is that we do not even have either mitochondria or eukaryotic cells in the first place.

- Why does the eukaryotic genome have a mosaic nature?
  The argument in this case is very similar to the previous one. The mechanist philosopher might try to isolate certain biochemical elements of the eukaryotic genome whose presence is responsible for the mosaicism. However, these different elements are merely biochemical and do not respond to the question asked in the first place; at most, the mechanist might provide us with a very good mechanistic explanation (in terms of parts, activities and arrangements) of why it is that the compounds of a eukaryotic genome admit mosaicism. But this does not explain why in fact all eukaryotes present this kind of genome and, more specifically, why they present the kind of additional genes they do with the particular functions they have. Again, we need a historical explanation to satisfy our queries and to ask "why" the genome in fact has a mosaic nature. As we said in Sect. 6.1, symbiogenesis can provide a reply to this: the fact that

distinct organisms came together to form the eukaryotic cell and have been living together for 1.5 billion years (with all the "arms races" that exist when distinct organisms live together) would explain this feature and would even explain the specific functions of the genes involved in such mosaicism (namely, these related to avoiding cheating on the part of mitochondria).

Once again, the mechanist philosopher might claim that, if, as biologists assume, the appeal to symbiogenesis provides the right answer, this is because symbiogenesis is a mechanism. But then the mechanist philosopher would have to acknowledge that the concept of a symbiogenetic process is so formal that no entities, activities or organization can be properly recognized and isolated, so as to identify a mechanism. Then the mechanist philosopher would have two options: either to relax the notion of mechanism, which would mean that the concept is made either empty or equivalent to the notion of regularity, or to accept that this fact is not explainable in mechanistic terms.

- Why are mitochondria phylogenetically closer to free-living alpha-proteobacteria than to their host?

  Here, the argument against mechanists precisely mimics that presented for question one, merely changing all the details concerning membranes for details concerning phylogenetic relations; so to save space, we will not repeat it.
- Why is there a gap in the fossil record between prokaryotes and eukaryotes?
  In this case, a defender of the mechanistic model of explanation might claim that we can always imagine a particular set-up with certain initial conditions and reason counterfactually (as Woodward proposes for historical explanations in general). Let us fix those set-up conditions. It would be a set-up where archaea and bacteria merge symbiotically. If we want to provide a reason why there is a gap, we have to isolate a factor such that, if we block it, the result would be different. Suppose for the sake of argument that symbiosis is such a factor and imagine that archaea evolve progressively until they give rise to eukaryotes.[5] Would this entail that there is no gap in the fossil record? Not necessarily. We have cases of gaps in the fossil record that are not due to symbiosis. For instance, nobody believes that the famous missing link between dinosaurs and birds is a consequence of symbiosis, despite this missing link creating a gap in the fossil record. Furthermore, there are examples of symbiotic mergers where no gap is present. Paracatenula is known to be the result of ancient endosymbiosis, but its existence does not entail that there is a gap in the fossil record between Paracatenula and other catenulid flatworms (Gruber-Vodicka et al. 2011). Therefore, reasoning counterfactually in the strict manner Woodward suggests does not help to explain this particular phenomenon. It seems that what is required is the assumption of a very particular pattern that follows a unique event (namely, a symbiotic merger). This pattern, due to

[5]Earlier in this section we argued, as part of the response to question 1, that it was quite hard to conceive of symbiosis as a factor. We still believe this, for the reasons discussed there, but we are going to assume here that it might serve as one, just for the sake of the argument.

> the complexity in determining its parts, activities and organization, cannot be interpreted mechanistically. In addition, it is difficult to see what a mechanistic reading in terms of parts, activities and organization can offer to explain the actual gap.

From this reasoning, three consequences follow. First, taking for granted that the appeal to symbiogenesis explains many of the issues about the origin of the eukaryotic cell, symbiogenesis is not, and it is very far from being, a mechanism.[6] Second, symbiogenesis seems to be more a general pattern which biologists appeal to in order to explain the features they find in the eukaryotic cell. Finally, even if the reference to mechanisms might complement explanations of some of the questions asked, and it might add some precision, the real explanatory role, as biologists accept, is played by the appeal to symbiogenesis.[7] Therefore, if symbiogenesis is not a mechanism but a general pattern, then it seems that the appeal to regularities might be explanatory after all.[8] In the next section, we further explore the possibility of considering symbiogenesis as a regularity.

6.4 Symbiogenesis as a Nomological-Expectable Explanation of the Origin of the Eukaryotic Cell

Biologists' appeal to the notion of symbiogenesis, as we have argued, has the form of a general pattern: the biologists look for a general principle, which may be quite vague (in the sense that it might be applicable to a large number of entities, irrespective of their particular biological properties), that allows them to say not only how the first eukaryotic cell came about, but also why it has the properties it has (which are the answers to the four why-questions we have presented, plus other

[6] In fact it would not even be a mechanism in the sense of "expanded mechanism" as defended by Roe and Baumgaertner (2016), since the problem here is not related to incorporating "pieces" of the environment, as they suggest: the problem is related to the fact that what plays the explanatory role is a regularity.

[7] One might still wonder about the exact relationship between mechanisms and regularities in certain explanatory contexts. It is not the aim of this paper to elucidate the nature of that relationship. Nonetheless, some mechanist philosophers have already recognized the use of non-accidental regularities in mechanistic explanations (e.g. Craver and Kaiser 2013; Glennan 1996) and we believe that, in most cases of mechanistic explanation, what does the real explanatory work is the presence of background non-accidental regularities. We plan to develop this line of thought further in a future paper.

[8] Of course, the defender of mechanistic explanation might still argue that the appeal to symbiogenesis is not, after all, explanatory. A similar strategy has been pursued by Craver (2008) concerning the explanatory character of the Hodgin-Huxley model of action potential in neurons. However, we believe that pursuing that strategy would violate some basic commitments common to biologists concerning explanation.

relevant questions that might be asked). It is convenient to specify at this point why we consider symbiogenesis to work as a regularity that might be used to account for certain facts (Archibald 2014; Douglas 2010).

First of all, symbiogenesis mere implies that the process by which an actual living organism has come about is a consequence of a symbiotic merger. Furthermore, in the case of the eukaryotic cell, it is always specified that this symbiogenesis gave rise to a case of endosymbiosis, whereby one organism lives inside the other. However, nothing about the particular nature of the organisms that interact endosymbiotically is specified, nor does it require to be specified in a general definition of symbiogenesis. Symbiogenesis just says something about how the mode of life of the organisms came about. Second, and related to the vagueness of the term, symbiogenesis is supposed to cover all the different cases of structures (and species) that emerge as a consequence of symbiosis between two different organisms. This entails that the entities that can interact symbiotically and give rise to a case of symbiogenesis are very different with respect to each other: bacteria, fungi, arthropods, mammals, etc.; they can all bear endosymbionts and/or enter endosymbiotic relationships with others. Third, by its very nature and its connection with the appearance of new biological structures, when it occurs through the acquisition of endosymbionts, symbiogenesis tends to trigger certain changes in the organisms involved: genomic decay, genetic assimilation, free exchange of genes between partners, vertical transmission, the appearance of particular bodily structures to bear the symbionts, etc. The evolution of these particular traits will differ depending on the particular relationship between the organisms and their necessities, and is normally what causes endosymbiotic states to be irreversible. Fourth and finally, symbiogenesis normally leaves some traces of the previously independent life of the partners. However, these traces vary quite a lot if we consider them on a case-by-case basis. Sometimes the traces will be biochemical pathways; others, molecular properties or chromosome structure, etc.

We believe that these four characteristics of symbiogenesis justify consideration of the phenomenon as a general pattern that biologists use in order to guide their research and to explain certain features that would not be explained otherwise. Indeed, the key aspect of symbiogenesis, in relation to accounting for the features of the eukaryotic cell as mentioned above, is that it makes these "expectable on the basis of [a] non-accidental regularit[y]" (Díez 2014, 1414). Nonetheless, this pattern, though general, is not empirically empty: it says something about the past and the future of the organisms which interact, and this can be studied further (and proved to be true or false). We believe that symbiogenesis, understood as we have specified above, is a kind of scientific law in Mitchell's sense (1997, 2000, 2003). In Mitchell's account, laws are understood pragmatically, according to the role they play in scientific practice. In other words, laws are not interpreted in terms of necessary and sufficient conditions, as traditional normative approaches suppose, but in terms of what they allow scientists to do. In this vein, Mitchell argues that a scientific statement must be understood as a scientific law if it allows good predictions to be made, good explanations to be provided and feasible interventions to be designed. This flexible conception of scientific laws allows her

to provide a new multidimensional framework to represent a whole set of scientific generalizations (Mitchell 2000, 259). Furthermore, scientific laws in this sense provide a certain non-zero degree of nomic necessity,[9] which is established in terms of the stability of the conditions upon which the regularity is contingently dependent.[10] Therefore, the degree of nomic necessity of regularities in physics is higher than that of regularities in biology, because the stability of the conditions upon which a regularity is contingent in physics and in biology are significantly different. However, both regularities in physic and in biology involve a certain degree of nomic necessity; which is what matters here and is relevant for considering these generalizations as legitimate scientific laws.

In the context of the symbiosis models of the origin of eukaryotes, the appeal to the concept of symbiogenesis seems to play the role of a scientific law in this sense. First, it is often supposed that endosymbiotic association between two different organisms will give rise to a tendency for a series of adaptations to evolve that will increase the tightness of the fit between the partners. These adaptations will tend to evolve due to the possible presence of "cheaters", i.e. organisms that benefit from the association without providing any benefit to its partner. This is a consequence of the fact that endosymbiotic associations that are capable of evolving adaptations that prevent the possible presence of cheaters outrun those that are not. Second, it is also assumed that the partners in an endosymbiotic association will still preserve some traces of their previous free-living state, as a consequence of the special features of the symbiogenetic process. Indeed, symbiogenesis sometimes entails (and it definitely does so in the eukaryote case) a transition in biological individuality. But, as is well known, the framework of transitions in individuality assumes the existence of individuals whose evolutionary fates align and form a higher-level entity. It is precisely the existence of independent individuals whose individualities become combined into a higher-level unit what makes it reasonable to expect that certain features of their previously independent existence will be preserved. In addition, the features that are preserved could be studied in a lab, making certain predictions possible. It is in at least these senses that we believe symbiogenesis plays the role of a nomic pattern (a pragmatic law): it allows for certain predictions, makes a set of phenomena that can be empirically tested expectable and supports counterfactuals. This nomic character seems to be the aspect of the notion of symbiogenesis that biologists have in mind when they use it for explanatory purposes.

Of course, defenders of mechanistic explanation might still question the alternative that we offer to mechanistic models of scientific explanation. As is well-known, the models that have traditionally appealed to scientific laws as the main explanatory "weapon" are conceptually flawed –they have to face numerous problems: flag-pole

[9]See also Brandon (1997) for more about biological generalizations having a limited range of nomic necessity and explanatory power.

[10]Mitchel also includes other parameters: ontological ones (strength, plus the aforementioned stability) and representational ones (degree of abstraction, simplicity and cognitive manageability), which we take not to be relevant for our purposes in this paper. See Mitchell (2003, chapter 5) for more details.

cases, contraception pills and male pregnancy, syphilis-paresis cases, vitamin C and flu recovery, etc.– and are not very popular among contemporary philosophers of science (Woodward 2017). Maybe, after all, we have to admit that, although not perfect, as our case illustrates, mechanistic explanation is the best theory of scientific explanation that we have for the moment. Nonetheless, Díez very recently proposed a new neo-Hempelian account that solves most of the conceptual problems that have been raised against nomological expectability models and –allegedly– would include mechanistic explanations as a specific subcase satisfying additional conditions (Díez 2002, 2014). As this is the only nomological alternative we know of that has these features, we now proceed to evaluate whether Díez's model can accommodate the case of the origin of the eukaryotic cell.[11]

Díez's model takes as a point of departure Hempel's thesis that "to explain a phenomenon is to *make it expectable* on the basis of non-accidental regularities" (Díez 2014, 1414). This expectability, however, is not as strict as it has traditionally been in deductive/inductive nomological models (one of the possible forms that nomological expectability models can take), where the cases in which the explanation is based on a low-probability relationship between the explanandum and the explanans were excluded. The reason for this exclusion was that explanations were taken as logical inferences; thus, in the case of inductive inferences, they demanded high probability (Hempel and Oppenheim 1948; Hempel 1965). In contrast, Díez substitutes the notion of logical inference for the less demanding notion of "embedding": according to Díez, to explain a phenomenon is to embed it "into some branch of a net of theoretical constraints" (the explanans) such that they make the phenomenon expectable (Díez 2014: 1419). The idea of embedding is the structuralist counterpart to the positivist notion of *implication* and it presupposes a distinction in scientific models/structures between *data models* and *theoretical models* (Balzer et al. 2012). A data model is a structure that describes the phenomenon to be explained; whereas theoretical models are the structures defined by their satisfying a theory's laws. A data model is embeddable in a theoretical model when the former "fits" the latter, i.e. the relevant values of the phenomenon square with those of the theoretical model. Embedding is thus a relation between models, not a relation between sentences, which allows for a weakening of the positivist demand for logical inference (for instance, making room for embedding in increasing yet law probability cases) but still preserves the core intuition behind Hempelian *expectability*. To put it in Díez's words "[e]xplanations are (at least) certain kinds of predictions" (Díez 2014, 1420).

We will now provide an example of embedding. Suppose we want to explain the movement of the Moon using Newtonian mechanics. Our data model would include the Earth, the Moon, and the space and time functions that describe the kinematic trajectory of the Moon around the Earth, [$DM_{E,M}$ = <{Earth, Moon}, space, time>]. The theoretical model would include, apart from the aforementioned components,

[11] What follows is mainly based on the analysis we already presented in Deulofeu and Suárez (2015).

the functions of *mass* and *force*, [$TM_{E,M}$ = <{Earth, Moon}, space, time, mass, force>] defined by their satisfying Newtonian laws. The idea of the embedding of the data model within the theoretical model would be the following: by using the "machinery" of classical mechanics (laws of motion) plus the relative positions of the Moon and the Earth at a particular time, the theoretical model includes the relevant positions at other times; if such values fit the measured values of the data model, the former successfully embeds the latter, otherwise the embedding fails (and the theory has a Kuhnian anomaly). In this sense, model-theoretical embedding expresses the core intuition of nomological expectability.

However, as Díez explains and the case of the Moon's trajectory exemplifies, nomological embedding, though necessary, is not sufficient for explanation, since we may still fail to have explanatory embedding in two kinds of cases. First, one may have embedding by merely descriptive/phenomenological theories that systematize data with laws describing general phenomena without explaining them (e.g. Galilean kinematics or Kepler's laws). Second, in theories with what Kuhn calls "general schematic principles" such as Newton's Second Law (Kuhn 1970), one can always construct *ad hoc* trivial "successful" embedding that cannot count as explanatory. To exclude these cases, Díez adds two further conditions: the embedding has to be *ampliative* and *specialized*. Its *ampliative* character is based on the notion of T-theoreticity (Balzer et al. 2012; related to Hempel's distinction between "characteristic" and "antecedently understood", and Lewis's distinction between old and new vocabulary). T-theoretical concepts are those introduced by a theory such that, in order to determine their extension, one has to use/accept some T-law (e.g. *mass* and *force* in classical mechanics); whereas T-non-theoretical concepts are those which are already available and that can be determined (at least on some occasions) without the help of T-laws (e.g. *space* and *time* in classical mechanics). Explanatory embedding is ampliative, as in the case of classical mechanics: classical mechanics explains why the Moon is in location X at time t through embedding the phenomenon and introducing new T-theoretical concepts/entities (masses and forces) that do not appear in the data model $DM_{E,M}$. Thus, for embedding to be explanatory, it must make use of laws that (as in classical mechanics and not in Galilean kinematics or Keplerian astronomy) appeal to new concepts/entities. *Specialization*, on the other hand, requires that we introduce non-*ad hoc* "special laws" in order to account for the phenomena.[12] As Díez points out, we always require that our explanations include something more than merely schematic, very general principles such as $\sum f = ma$. In the case of the Moon–Earth system, for example, we need to introduce the law of universal gravitation, $f = G*mm'/r^2$, if we aim to explain the positions of the Moon over time.

In short, we might now say that a theory explains a phenomenon if: (1) we can embed the phenomenon in the theory, in such a way that the theory makes the phenomenon expectable; (2) the theory includes and makes use of at least one T-theoretical term; and (3) the theory incorporates and makes use of at least one special

[12] As we said before, the notion of law that we use is Mitchell's idea of pragmatic law.

law in order to account for the phenomenon (Díez 2014, 1425). We will show that the appeal to symbiogenesis that biological theory makes to explain the origin of eukaryotes and the different phenomena laid out in Sect. 6.1, which does not fit the mechanistic account, is nevertheless perfectly legitimate and can be explicated by applying Díez's re-elaborated model of explanation as nomological expectability.

First, the appeal to symbiogenesis provides a theoretical model that allows the embedding of the phenomena that constitute our data model. In the case of the origin of the eukaryotic cell, the data model would include objects such as membranes –of both cells and mitochondria– or genomes –again, both cell and mitochondrial genomes– and their respective biochemical properties –those of the lipid components of the mitochondrial membrane versus those of the lipid components of the cell membrane; circular, single-strand DNA versus linear, complex DNA, etc.– ($DM_{G,M}$ = <{genome, membrane}, biochemical properties of both>). The theoretical model would include these objects plus entities/functions that correspond to the notions of *fitness* and *symbiogenesis*, which are purely theoretical and associated with particular features of symbiosis relationships and the theory of natural selection ($TM_{G,M}$ = <{genome, membrane}, biochemical properties of both, fitness, symbiogenesis>).[13] The embedding is possible in this case because $DM_{C,M}$ happens to actually be a submodel that squares with $TM_{G,M}$, and $TM_{G,M}$ makes the phenomena we aim to explain expectable (as reviewed in Sect. 6.1 in response to questions 1–4).

Furthermore, $TM_{G,M}$ includes a couple of T-theoretical entities/functions, fitness and symbiogenesis, that play an *ampliative* role. Biologists do not explain the features of the mitochondrial genome by appealing to features of free-living bacteria. They explain them by appealing to the idea of symbiogenesis (and its specific endosymbiotic form): certain formerly free-living bacteria (that we can indicate through phylogenetic analysis) were at some point endosymbiotically acquired by an archaeon and, *symbiogenetically*, gave rise to the organelles that nowadays we call mitochondria. The preservation of the genetic and biochemical features of the mitochondrial membrane is explained by appealing to its symbiogenetic origin plus the fact that they confer fitness advantages. In this sense, it seems clear that the embedding is ampliative in the sense Díez's account requires.

Finally, the explanation in terms of symbiogenesis includes an element of specialization in relation to ST (or the concept of symbiosis): an appeal to a special law which plays a non-trivial role in the explanation of the particular features of mitochondria. Symbiogenesis is a particular form of integration that two symbiotically associated organisms could enact, if the circumstances were favourable. It is well established that there are different types of symbiotic relationship (mutualism, commensalism and parasitism); some might be long-term evolutionarily relationships that are not conducive to integration, whereas others

[13]This reconstruction is merely informal and, due to particular complexities of biological theory, it cannot be made as precise as it could be in the case of classical mechanics. In any case, it has all the elements that are supposed to provide a general idea concerning embedding.

are. If they are conducive to integration and they have the desired fitness effects (i.e. they do not lead to the extinction of the integrated lineages), then they would trigger certain changes in the lineages that evolve symbiogenetically (mosaicism, genomic decay, loss of independent modes of life, etc.), giving rise to the appearance of new biological structures (they would fall down an evolutionary "rabbit hole", as some biologists describe it, e.g. Moran and Sloan 2015). In contrast, if the symbiosis relationship does not lead to integration, even if it is a long-term relationship, it would lead to a different kind of changes that would affect to both organisms independently, such as certain phenotype changes, changes in behaviour, etc. In this sense, symbiogenesis plays the role of a special law concerning a more general principle of the expected outcomes of long-term symbiotic associations.

We believe this appeal to a special law is the crucial step in ST, it is what provides the main explanatory power and as we argued, it does not have the form of a mechanism. The special symbiosis law certainly is such in Mitchell's pragmatic sense: it provides a certain degree of nomic necessity, therefore providing biologists with a guide to what they might find. For instance, appealing to a symbiogenetic origin makes it expectable that organelles, i.e. mitochondria, still preserve a certain degree of biological individuality that might be manifested, for example, by the possibility of in vivo replication. It is important to bear in mind that this would not be expected if the origin was self-genetic: in this latter scenario, we would never expect mitochondria to have any degree of biological individuality. Furthermore, if the origin of mitochondria is symbiotic, we will not expect to find intermediate forms in the fossil record, since symbiosis gives rise to saltational evolutionary events, which would not be the case if the origin was self-genetic. This same line of reasoning might be applied to all the features that ST makes nomically expectable and, in this sense, we have something similar to a pragmatic law that provides the research field with some order.

We should still note something about the special character of the law. As we said before, the condition is introduced in order to avoid counting as explanatory cases in which we merely apply general principles to *trivially* justify why certain phenomena occur (using *ad hoc* mathematical functions in $\sum f = ma$ to explain intentional movement, for instance). One might argue that the appeal to symbiogenesis is still trivial in this last sense: it is just one general principle we could use to justify every feature we find in an organism. Nonetheless, this is not the case: the appeal to symbiogenesis rules out certain possibilities and it makes a difference (as does the appeal to $f = G*mm'/r^2$, in the case of planetary movement). It specifies the manner in which evolutionary innovation can arise, and this is in contrast to other possibilities, such as mutation, recombination, methylation, changes in the developmental matrix, or even other types of long-term non-integrative symbiotic relationships. It specifies a very particular pattern followed by the organisms that experience this mode of generation of evolutionary novelties and precludes triviality by ruling out the appearance of certain features that other evolutionary pathways would make expectable.

In conclusion, we have provided a (partially weakened, partially strengthened) nomological expectability framework as a possible alternative to a mechanistic

framework of scientific explanation that explicates why biologists consider ST a legitimate explanatory theory of the origin of the eukaryotic cell by appealing to the notion of (pragmatic) scientific laws. In this sense, we have provided reasons to justify why an account of scientific explanation in terms of laws (in the restricted sense we have given) might be appealing to gain an understanding of the explanatory practices of biologists in certain contexts; an understanding that –we have claimed (Sect. 6.3)– mechanist philosophers cannot provide.

6.5 Concluding Remarks

In this paper we have presented the symbiosis model of the origin of the eukaryotic cell together with a set of questions (phylogenetic, biochemical, etc.) that any theory of the origin of the eukaryotic cell must provide answers to. We argue that the notion of symbiogenesis, understood as the process by which a new biological structure (organ, metabolic pathway, etc.) originates as a consequence of a long-term symbiotic relationship, plays the entire explanatory role when biologists aim to provide an answer to the different questions we mention (Sect. 6.2). This said, we defend the idea that the mechanistic account of scientific explanation is not well-suited to understanding why the notion of symbiogenesis plays the entire explanatory role in these cases. First, we argue that every attempt to offer a mechanistic explanation to the questions previously mentioned turns out to be unsatisfactory, since they move to a level of detail which turns out to be unnecessary for the matters discussed; moreover, many of the causes that should be mentioned in a mechanistic account seem orthogonal to the type of phenomena that demands an explanation. Second, we show that the notion of symbiogenesis is far from being a mechanism as they are conventionally understood in the literature (in terms of parts, activities and organization): symbiogenesis is a regularity or general pattern that cannot be suitably captured in mechanistic terms (Sect. 6.3). Finally, we present Díez's nomological expectability model of scientific explanation as an alternative to mechanistic models of explanation and defend the notion that Díez's model helps in understanding the explanatory character of symbiogenesis, despite its not being a mechanism but a general pattern (Sect. 6.4). If our argument is sound, it shows how and why the appeal to general patterns –that might well be considered scientific laws in Mitchell's sense, as we argue– might be explanatory in some contexts, thus challenging the universality of mechanistic explanations. It remains to be explored, however, whether the nomological expectability approach to scientific explanation we have defended here could also be applied to other biological contexts, either as a complement to (e.g. Alleva et al. 2017) or as a substitute for mechanistic accounts.

Acknowledgments Different versions of this paper were presented at the VIII Meeting of the Spanish Society for Logic, Methodology and Philosophy of Science (University of Barcelona, 2015) and the III Conference of the German Society for the Philosophy of Science (University of Düsseldorf, 2016). We would like to thank all the participants for their helpful comments and

suggestions. We would also like to thank Thomas Bonnin, Mark Canciani, José Díez, John Dupré, Çağlar Karaca, Adrian Stencel and an anonymous referee, who read previous versions of this paper and made helpful comments and suggestions. Finally, the Spanish Ministry of Economy and Competitiveness (FFI2016-767999-P) and the Fundación Bancaria la Caixa are formally acknowledged for their economic support.

References

Alleva, K., J. Díez, and L. Federico. 2017. Models, theory structure and mechanisms in biochemistry: The case of allosterism. *Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences* 63: 1–14.

Archibald, J. 2014. *One plus one equals one: Symbiosis and the evolution of complex life*. Oxford: Oxford University Press.

———. 2015. Endosymbiosis and eukaryotic cell evolution. *Current Biology* 25: R911–R921.

Audesirk, T., G. Audesirk, and B.E. Byers. 2008. *Biology: Life on earth*. Hoboken: Pearson Education.

Balzer, W., C.U. Moulines, and J.D. Sneed. 2012. *An architectonic for science: The structuralist program*. Vol. 186. Dordrecht: Springer.

Bechtel, W. 2006. *Discovering cell mechanisms: The creation of modern cell biology*. Cambridge University Press.

Bechtel, W. 2011. Mechanism and biological explanation. *Philosophy of Science* 78 (4): 533–557.

Bechtel, W., and A. Abrahamsen. 2005. Explanation: A mechanist alternative. *Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences* 36 (2): 421–441.

Bechtel, W., and R.C. Richardson. 1993. *Discovering complexity: Decomposition and localization as scientific research strategies*. Cambridge, MA: The MIT Press.

Brandon, R.N. 1997. Does biology have laws? The experimental evidence. *Philosophy of Science* 64: S444–S457.

Cavalier-Smith, T. 1989. Archaebacteria and Archezoa. *Nature* 339: 100–101.

———. 2013. Symbiogenesis: Mechanisms, evolutionary consequences, and systematic implications. *Annual Review of Ecology, Evolution, and Systematics* 44: 145–172.

Craver, C.F. 2006. When mechanistic models explain. *Synthese* 153 (3): 355–376.

———. 2007. *Explaining the brain*. New York: Clarendon Press.

———. 2008. Physical law and mechanistic explanation in the Hodgkin and Huxley model of the action potential. *Philosophy of Science* 75: 1022–1033.

Craver, C.F., and M.I. Kaiser. 2013. Mechanisms and laws: Clarifying the debate. In *Mechanism and causality in biology and economics*, ed. H.K. Chao, S.T. Chen, and R.L. Millstein, vol. 3, 125–145. New York: Springer.

Darden, L., and C.F. Craver. 2002. Strategies in the interfield discovery of the mechanism of protein synthesis. *Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences* 33 (1): 1–28.

Deulofeu, R., and J. Suárez. 2015. On the explanatory character of Margulis' theory of the origin of the eukaryotic cell. In *Proceedings of the VIII Meeting of the Spanish society for logic, methodology and philosophy of science*, ed. J.A. Díez, M. García-Carpintero, J. Martínez, and S. Oms, 276–280. Barcelona: Universitat de Barcelona.

Díez, J.A. 2002. Explicación, unificación y subsunción. In *Diversidad de la explicación científica*, ed. W.J. González, 73–96. Barcelona: Ariel.

———. 2014. Scientific w-explanation as ampliative, specialised embedding: a neo-hempelian account. *Erkenntnis* 79: 1413–1443.

Dolan, M.F. 2013. Endosymbiosis and the origin of eukaryotes. In *Evolution from the Galapagos: Two centuries after Darwin*, ed. G. Trueba and C. Montúfar, 125–128. New York: Springer.

Douglas, A.E. 2010. *The symbiotic habit*. Oxford: Princeton University Press.
Glennan, S. 1996. Mechanisms and the nature of causation. *Erkenntnis* 44 (1): 49–71.
———. 2002. Rethinking mechanistic explanation. *Philosophy of Science* 69 (S3): S342–S353.
Gruber-Vodicka, H.R., U. Dirks, N. Leisch, C. Baranyi, K. Stoecker, S. Bulgheresi, N.R. Heindl, M. Horn, C. Lott, A. Loy, M. Wagner, and J. Ott. 2011. Paracatenula, an ancient symbiosis between thiotrophic Alphaproteobacteria and catenulid flatworms. *PNAS of the United States of America* 108 (29): 12078–12083.
Hempel, C.G. 1965. *Aspects of scientific explanation and other essays in the philosophy of science*. New York: Free Press.
Hempel, C., and P. Oppenheim. 1948. Studies in the logic of explanation. *Philosophy of Science* 15 (2): 135–175.
Kuhn, T.S. 1970. Second thoughts on paradigms. In *The structure of scientific theories*, ed. F. Suppe, 459–482. Urbana: University of Illinois Press.
Leuridan, B. 2010. Can mechanisms really replace laws of nature? *Philosophy of Science* 77 (3): 317–340.
Machamer, P., L. Darden, and C.F. Craver. 2000. Thinking about mechanisms. *Philosophy of Science* 67 (1): 1–25.
Margulis, L. 1970. *Origin of eukaryotic cells: Evidence and research implications*. New Haven: Yale University Press.
Martin, W.F., M. Roettger, T. Kloesges, T. Thiergart, C. Woehle, S. Gould, and T. Dagan. 2012. Modern endosymbiotic theory: Getting lateral gene transfer into the equation. *Journal of Ensocytobiosis and Cell Research* 23: 1–5.
Martin, W.F., S. Garg, and V. Zimorski. 2015. Endosymbiotic theories for eukaryote origin. *Philosophical Transactions of the Royal Society B* 370: 20140330.
Mitchell, S.D. 1997. Pragmatic laws. *Philosophy of Science* 64: S468–S479.
———. 2000. Dimensions of scientific law. *Philosophy of Science* 67 (2): 242–265.
———. 2003. *Biological complexity and integrative pluralism*. Cambridge: Cambridge University Press.
Moran, N., and D.B. Sloan. 2015. The hologenome concept: Helpful or hollow? *PLoS Biology* 13 (12): e1002311.
Nicholson, D.J. 2012. The concept of mechanism in biology. *Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences* 43 (1): 152–163.
O'Malley, M. 2010. The first eukaryote cell: An unfinished history of contestation. *Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences* 41 (3): 212–224.
Raff, R.A., and H.R. Mahler. 1972. The non symbiotic origin of mitochondria. *Science* 177: 575–582.
Roe, S.M., and B. Baumgaertner 2016. Extended mechanistic explanations: Expanding the current mechanistic conception to include more complex biological systems. *Journal for General Philosophy of Science* 48: 517–534
Sagan, L. 1967. On the origin of mitosing cells. *Journal of Theoretical Biology* 14: 225–274.
Sapp, J. 2010. Saltational symbiosis. *Theory Biosciences* 129: 125–133.
Spang, A., J.H. Saw, S.L. Jørgensen, K. Zaremba-Niedzwiedzka, J. Martijn, A.E. Lind, R. van Eijk, C. Schleper, L. Guy, and T.J.G. Ettema. 2015. Complex Archaea that bridge the gap between prokaryotes and eukaryotes. *Nature* 521: 173–179.
Stearns, S.C., and R.F. Hoekstra. 2000. *Evolution. An introduction*. Oxford: Oxford University Press.
Uzzel, T., and C. Spolsky. 1974. Mitochondria and plastids as endosymbionts: A revival of special creation? *American Scientist* 62: 334–343.
Williams, T.A., and T.M. Embley. 2015. Changing ideas about eukaryotic origins. *Philosophical transactions of the Royal Society of London. Series B* 370: 20140318.
Woodward, J. 1997. Explanation, invariance, and intervention. *Philosophy of Science* 64: S26–S41.

———. 2000. Explanation and invariance in the special sciences. *The British Journal for the Philosophy of Science* 51 (2): 197–254.
———. 2003. *Making things happen*. New York: Oxford University Press.
———. 2017. Scientific explanation. In The *Stanford encyclopedia of philosophy*, ed. E.N. Zalta. https://plato.stanford.edu/archives/spr2017/entries/scientific-explanation/

Synthese
https://doi.org/10.1007/s11229-019-02187-9

BIOBEHAVIOUR

Explaining the behaviour of random ecological networks: the stability of the microbiome as a case of integrative pluralism

Roger Deulofeu[1] · **Javier Suárez**[1,2] · **Alberto Pérez-Cervera**[3]

Received: 12 June 2018 / Accepted: 20 March 2019

Abstract

Explaining the behaviour of ecosystems is one of the key challenges for the biological sciences. Since 2000, new-mechanism has been the main model to account for the nature of scientific explanation in biology. The universality of the new-mechanist view in biology has been however put into question due to the existence of explanations that account for some biological phenomena in terms of their mathematical properties (mathematical explanations). Supporters of mathematical explanation have argued that the explanation of the behaviour of ecosystems is usually provided in terms of their mathematical properties, and not in mechanistic terms. They have intensively studied the explanation of the properties of ecosystems that behave following the rules of a non-random network. However, no attention has been devoted to the study of the nature of the explanation in those that form a random network. In this paper, we cover that gap by analysing the explanation of the stability behaviour of the microbiome recently elaborated by Coyte and colleagues, to determine whether it fits with the model of explanation suggested by the new-mechanists or by the defenders of mathematical explanation. Our analysis of this case study supports three theses: (1) that the explanation is not given solely in terms of mechanisms, as the new-mechanists understand the concept; (2) that the mathematical properties that describe the system play an essential explanatory role, but they do not exhaust the explanation; (3) that a non-previously identified appeal to the type of interactions that the entities in the network can exhibit, as well as their abundance, is also necessary for Coyte and colleagues' account to be fully explanatory. From the combination of these three theses we argue for the necessity of an integrative pluralist view of the nature of behaviour explanation when this is given by appealing to the existence of a random network.

The paper is the result of the discussion among the three authors, who actively collaborated in the development of all the ideas. JS conceived and structured it. RD and APC wrote Section 3. RD and JS wrote the philosophical analysis.

✉ Roger Deulofeu
Roger.deulofeu@gmail.com

Extended author information available on the last page of the article

Published online: 03 May 2019

Keywords Scientific explanation · Mechanism · Mathematical explanation · Behaviour explanation · Integrative pluralism · Random network

1 Introduction

Explaining the behaviour of ecosystems is one of the key challenges for biologists: why ecosystems have the properties they have, which conditions make them exhibit a stable behaviour, how they react to perturbations, etc. have been some of the most debated questions among behavioural ecologists. Almost since the original publication of the "new-mechanist" *manifesto* (Machamer et al. 2000), the idea that explanation in biology proceeds by discovering mechanisms has been extensively accepted among philosophers of science (Bechtel and Richardson 1993; Glennan 2002; Bechtel and Abrahamsen 2005; Craver and Darden 2005, 2013; Craver 2007). Drawing upon the findings in neuroscience and molecular biology, new-mechanists reelaborate causalism (Salmon 1984; Woodward 2003) and argue that to explain a phenomenon (*explanandum*) consists in describing its causes by providing a mechanism (*explanans*) responsible for the phenomenon coming about. According to the new-mechanist interpretation of explanation, a phenomenon can be considered fully explained only once a mechanism is provided and the way its components causally interact to produce the phenomenon is specified.

The universality of mechanistic explanations in biology and neuroscience as originally formulated in the new-mechanist *manifesto* has however been questioned on different grounds. A very popular "non"-mechanist view holds that some explanations in biology are at least *partially* given by appeal to the mathematical properties of the systems under investigation. Some of those who defend the necessity of introducing mathematical modelling in some biological explanations have emphasized the continuity of these explanation with mechanistic explanations, developing the concept of *dynamic mechanistic explanations* (Bechtel and Abrahamsen 2010; Bechtel 2011; Brigandt 2013a, b, 2015), whereas others have highlighted the differences between both accounts, arguing that dynamic mechanistic explanation are indeed no-mechanistic (Issad and Malaterre 2015). A still more radical interpretation of this "non"-mechanist approach holds that some explanations in biology are given exclusively in mathematical terms, with no reference to any type of causal-mechanistic information in the *explanans* (Sober 1983; Huneman 2010; Lange 2013; Jones 2014). This last kind of explanation might be called *mathematical explanation*, and it is usually presented in complete opposition to new-mechanist models of explanation.

One of the most recent topics of attention of defenders of the last approach has been the study of the explanation of the behaviour of ecosystems. Some of the most salient behavioural properties of ecosystems (robustness, resilience, stability), they argue, are explained using the tools provided by network analysis (Xia 2010; Rathkopf 2018). The use of these mathematical tools to explain how some biological systems behave in the long-term is very abundant in ecology (e.g. the diversity-stability debate) due to the complexity of ecological systems. It has recently been argued that explanations of the behaviour of ecosystems are given exclusively on the basis of the mathematical properties of the networks that are used to represent them (Huneman 2010, 2018a, b,

c). The study of behaviour explanations that appeal to network analysis in ecology has focused exclusively on the study of ecosystems that form *non-random networks*, though. However, as far as we know, the explanatory features of behaviour explanations in ecosystems that organize forming *random networks* have received no attention among scholars. A network is said to be non-random when the aggregation of the elements that interact in the network lead to a concrete topological realization (small world, scale-free, etc.) with some a priori known properties. A random network, on the contrary, is characterized for lacking a known topological realization, and thus the properties of a random network cannot be known a priori, but have to be mathematically discovered. We suspect that the difference between systems that organise according to a random network and systems that organise according to a non-random network get their explanatory force from different sources. In this paper, we aim to test that intuition by studying the explanatory features of the models that explain the ecological behaviour of a random network—the human microbiome.

A microbiome is a collection of microorganisms (bacteria, viruses, fungi, etc.) that reside in a concrete environment (Marchesi and Ravel 2015). In case of humans, our microbiome is believed to contain hundreds of species (Human Microbiome Project Consortium 2012). Furthermore, human's gut microbiome is known for its ecological stability behaviour: even if it varies quite a lot from one person to another, the gut microbiome is believed to be very stable for one single individual, who tends to carry the same species of microbes for a long period (Dethlefsen and Relman 2011; Faith et al. 2013). Recently, Coyte et al. (2015) and Foster et al. (2017) have elaborated a model to explain why the human gut microbiome exhibits a stable behaviour over long periods of time. They found out that, contrary to a usual assumption in evolutionary biology, competition, and not cooperation, is the key factor explaining this stable behaviour. Here, we analyse how Coyte and colleagues explain the stability of the microbiome by appealing to linear stability analysis. This case shares several elements with other cases of behaviour explanation in terms of networks, although it also differs in some features that led to some differences regarding the nature of behaviour explanation and that we aim at clarifying.

In Sect. 2 we introduce the notions of mechanistic and mathematical explanation, and frame the choice of our case study in the context of the mechanistic versus mathematical debate. In Sect. 3 we present Coyte and colleagues' explanation of the stability behaviour of the human microbiome. In Sects. 4 and 5 we discuss the consequences of the case study for the understanding of scientific explanation. In Sect. 4 we argue that even if the explanation in our case study displays a model of mechanism, it does not specify a proper causal story to account for the *explanandum*, thus suggesting that mechanisms, as understood by new-mechanists, do not play the explanatory role. In Sect. 5 we argue that the explanatory force in Coyte and colleagues' account comes from the display of a mathematical model of the behaviour of the microbiome, as it is provided by their linear stability analysis of the network that the microbiome instantiates. However, in contrast with some recent analysis of behaviour explanations of ecosystems in terms of networks, we argue that insofar as an essential component of the explanatory force of Coyte et al.'s model is their appeal to the different interaction types within the ecological communities that might exist in the microbiome (cooperative, exploitative, competitive), and thus to the establishment of a range of topologies—and

not exclusively to a concrete topology—their explanations differs from purely mathematical explanations. We further argue that the appeal to this element is exclusive of random networks. In Sect. 6 we explore the pluralistic consequences of our case study for the analysis of behaviour explanation, suggesting that it supports the general appeal to integrative pluralism. Finally, we present our concluding remarks.

2 Two models of explanation in biology: mechanistic versus mathematical

In its most basic meaning, to explain a phenomenon consists in giving the reasons why the phenomenon obtains in a concrete system due to its behaviour. Those reasons may be specified by appealing to laws of nature—deductive-nomological theories of explanation—or by appealing to the causes that are responsible for the phenomenon. In the first case, it is usually assumed that to explain a phenomenon is to provide an argument so that the phenomenon (*explanandum*) logically follows from the laws of nature that regulate the behaviour of the system where it obtains, given a concrete set of initial conditions (*explanans*) (Díez 2014; Alleva et al. 2017). In the second case, the phenomenon is embedded in a causal network in the world, in a way such that the phenomenon (*explanandum*) is a causal consequence of the behaviour of the system (*explanans*) (Woodward 2017).

New-mechanist theories of scientific explanation are of this last kind. They assume that to explain a phenomenon consist in citing its causes by providing a mechanism that specifies how those causes produce the phenomenon. The definition of "mechanism" is different in differing accounts (Nicholson 2012; Deulofeu and Suárez 2018), although most new-mechanists share a similar conception: a mechanism consists in a set of *entities* with a concrete spatial *organization* plus a set of *activities* governing the behaviour of those entities (*model of the mechanism*). The phenomenon to explain, new-mechanists argue, *causally* obtains as a consequence of the activities of the entities (*causal story*) (e.g. Machamer et al. 2000; Bechtel and Abrahamsen 2005; Craver and Darden 2005 2013; Craver 2007; etc.). For new-mechanists, the presence of a *model of the mechanism* and of a *causal story* is necessary and sufficient for having a mechanistic explanation (Issad and Malaterre 2015: p. 270).

New-mechanists generally accept a hierarchical view of mechanisms, thus not confining themselves to the narrow approaches of previous causalists (Salmon 1984). Furthermore, they usually neglect the capacity of mathematical models *alone* to explain any phenomenon. In a well-known paper, Kaplan and Craver argued that "the [mathematical] generalizations are explanatory because they describe the causal relationships that produce, underlie, or maintain the *explanandum* phenomenon" (2011: p. 612) and insisted that:

> In successful explanatory models in cognitive and systems neuroscience (a) the variables in the model correspond to components, activities, properties, and organizational features of the target mechanism that produces, maintains, or underlies the phenomenon, and (b) the (perhaps mathematical) dependencies posited among these variables in the model correspond to the (perhaps quantifi-

able) causal relations among the components of the target mechanism. (2011: p. 611).

Mathematics, thus, can play an explanatory role for new-mechanists only if it captures *a causal relationship* among the entities that are posited in the mechanism. Otherwise, they are merely "phenomenological models", which represent the reality and allow predictions without really explaining why those predictions obtain (e.g. Kepler's laws, Snell's laws, etc.) (Díez 2014).

Radically contrasting with this last view, some people have vindicated a more substantial role for mathematics in biological explanations by highlighting the importance of mathematical properties for explaining the features of some biological systems. The explanations that rely on the mathematical properties of the system to explain a phenomenon have been called "mathematical explanations" (Baker 2015), or "structural explanations" (Huneman 2018a).[1] They have been defined as follows:

> Family of explanations for which the mathematical tools used in the description of an explanandum system belong to a mathematical structure whose properties are directly explanatory of some aspects of the system (such as equilibria, behaviour, limit regime, asymptotic behaviour, etc.) (…) They explain by accounting for the explananda through pinpointing structural relations that are mathematical relations of some sort. (Huneman 2018a: p. 695)

The mathematical properties that appear in a mathematical explanation might be of different types, and they could be used to explain different kinds biological questions. They might consist in: the application of an arithmetic theorem to explain the life cycle of some species (Baker 2005, 2009, 2015); the establishment of one or more points of equilibrium to explain a tendency in a population (Sober 1983; Kuorikoski 2007; Rice 2012, 2015; Suárez and Deulofeu, unpublished manuscript); the application of statistics to explain certain evolutionary patterns in a population (Walsh 2015); the discovery of a concrete topology that explains the behaviour of a complex system (Huneman 2010; Jones 2014); the use of matrix calculus to explain the processes that regulate some physiological states (Issad and Malaterre 2015); etc.[2]

Not every substantial use of mathematics in biological explanations needs to be in principle completely opposed to every element of the new-mechanistic account of explanation, though. The explanation of some cyclical biological processes such as the circadian rhythms has been argued to constitute an extension of mechanistic explanation, namely a *dynamic mechanistic explanation* (Bechtel and Abrahamsen 2010, 2011; Brigandt 2013a, b, 2015). In general, an explanation is considered a dynamic mechanistic explanation in virtue of making use of some mathematical model in its *explanans* that: (1) is essential to account for the *explanandum*, (2) replaces the role that new-mechanists attribute to the causal story, without being itself a causal

[1] To refer to mathematical explanations as "structural explanations" might be confusing, since the later could be interpreted as special cases of the former, as one reviewer has correctly suggested. However, the way in which Huneman (2018a) describes them, as well as the family of explanations that he includes under the umbrella of "structural explanations" makes clear that the two are synonymous. For purposes of clarity, however, we will refer to this family of explanations as "mathematical explanations".

[2] We take all the aforementioned properties to be different types of mathematical properties.

story *stricto* sensu, (3) is combined with a model of mechanism, i.e. a set of entities and activities plus their organization. Nonetheless, it is precisely because of the lack of a causal story that some of these explanations entail that they have been argued to be "anti"-mechanistic, or at least far away from the core elements that new-mechanists consider necessary and sufficient to formulate an adequate scientific explanation (Issad and Malaterre 2015).

In still more extreme cases, however, some explanations in biology have been argued to be even more substantially mathematical, abstracting away also from the model of the mechanism, and being explanatory of the biological phenomenon *exclusively* in virtue of the mathematical properties of the *explanans*, whatever their type (Sober 1983; Baker 2005; Huneman 2010; Jones 2014; cf. Kuorikoski 2007; Potochnik 2015). The analysis of behaviour explanations in biological systems that present a network structure (e.g. ecosystems, immunological systems, etc.) has been argued to follow this pattern of mathematical explanation. The behaviour of this type of systems is usually explained in two steps: (1) the system is attributed a concrete network topology, which provides the mathematical properties of the system; (2) the properties of this topology are studied and then its behaviour is attributed to the biological system, explaining why the system behaves how it does. Because of the reliance of the last type of mathematical explanation on the topological properties of the networks, it has been called *topological explanation* (Huneman 2010, 2018b; Jones 2014; Brigandt et al. 2017).

Those who have studied topological explanation more attentively have made two points: First, that the appeal to the topological properties of the mathematical structure alone (its graph structure, or its network motifs, for instance) is *sufficient* to explain some of the properties of the biological system that the structure is believed to represent, irrespectively of the entities and the activities of the entities that realize those systems (see also Huneman 2018c); second, that the addition of any mechanistic details, instead of making the explanation of the properties more precise, obfuscates the question and turns out to be deeply irrelevant for the embedding of the *explanandum*. This second point might be explicated as follows: in sharp contrast with dynamic mechanistic explanations, in topological explanations, neither the model of mechanism (nature of the entities, nature of the activities), nor the causal story are explanatory relevant. All that matters in the explanation is that the topology is provided and that it gives information about the organization of the system.

One of the fields where the appeal to topological properties to explain biological phenomena has proven more fruitful is in the diversity-stability debate in ecology (McCann 2000; Nikisianis and Stamou 2016). In that context, the aim of ecologists is to elaborate network models that represent the relationships among the biotic members of an ecological community with the aim of inferring some general features about its behavioural patterns. To do so, once the network model is elaborated and linked to behavioural properties of the ecological community which is being studied, ecologists analyse the global properties of the network—e.g. how it will respond to a perturbation, to an increasing/decreasing number of nodes, to an increase in the number of connections, etc., and then attribute the exact same properties to the ecological community that the network is believed to capture. The driving question in the diversity-stability debate is whether increasing the number of species in the community would make the

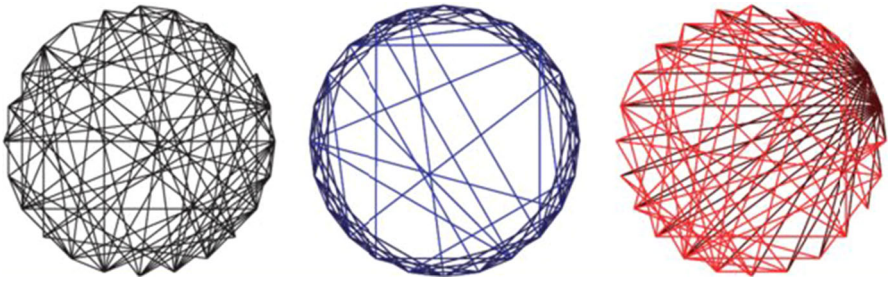

Fig. 1 This figure exemplifies different types of networks. The one on the left is a random network. The one in the middle is a small world network. The one of the right is a scale free network. (From Sporns et al. 2004: p. 419)

community ecologically stable[3] and, if so, under which conditions. Some recent discussion surrounding the diversity-stability hypothesis have tried to unravel how some ecological communities will react to the loss of some species for communities where the connections between the nodes are non-random (Solé and Montoya 2001). In the context of network theory, a network is non-random if it instantiates a particular topology, e.g. scale-free networks, small worlds (Fig. 1). Because the topological properties of theses types of networks are known, and both small worlds and scale-free networks are known to be highly stable to the elimination of some of their nodes (Montoya and Solé 2002), it is *enough* for ecologists to prove that a concrete ecosystem instantiates one of these networks to explain why the ecosystem exhibits a stable behaviour. The explanation in these cases would work as follows:

> Ecosystem *E* instantiates a network *N* which, in virtue of being of type *X* has the topological property *P*. Therefore, *E* also has *P* (adapted from Huneman 2010).

Interestingly, these types of explanations: (1) do not mention either the entities or the activities that might be going on in the ecosystem, insofar as network analysis only represents relations in terms of the number of nodes and the strength of their interactions—thus being applicable to multiple kinds of systems, just replacing "node" for the objects that are studied in the field (Internet, metabolic networks, social networks, etc.); (2) do not elaborate any kind of causal story that is responsible of producing the phenomenon under investigation. The *explanandum* (*E* having *P*) is accounted for simply because the network is of type *X*, and thus necessarily must instantiate P. Thus, topological properties alone would explain *P* obtaining, and there is no role left for mechanisms (Huneman 2010).

The study of behaviour explanations provided in terms of networks analysis has been centred in the study of explanatory patterns in systems that instantiate non-random networks (e.g. scale-free networks, or small worlds). However, no attention has been devoted to study the explanatory patterns that underlie behaviour explana-

[3] The exact definition of stability is an agitated topic in ecology, and different diversity-stability hypotheses are formulated accordingly (McCann 2000: p. 230, Table 1; Nikisianis and Stamou 2016: pp. 35–36; Gonze et al. 2018: p. 42, Box 1). In most cases, though, a system is qualified as stable if and only if it is able to return to its initial state after a perturbation (resilience), or also the capacity of a population to resist invasions by external species. We will specify later what "stability" means in our case study.

tions for systems that instantiate purely random networks. We suspect that because random networks neither have a particular topology associated, nor an exclusive type of network motifs, the type of explanatory patterns that underlie behaviour explanations of the systems that instantiate them might be slightly different from the types of explanatory patterns provided in terms of non-random networks. The rest of the paper is thus devoted to analyse the explanatory features of a behaviour explanation given for a system that instantiates a random network.

3 Case study: the stability of the human microbiome

The human gut is an ecosystem consisting in a large community of microbes (≈ 1000 species), whose stability behaviour is crucial to maintain human's health.[4] Recent empirical research suggests that the human microbiome exhibits a stable behaviour: even if different individuals might bear different microorganisms in their microbiome, the species that compose the microbiome of an individual, and their relative densities, tend to remain largely stable during her lifetime (Dethlefsen and Relman 2011; Faith et al. 2013). The reasons that make such essential community to behave stably despite the existence of constant perturbations are yet unknown, though. One possible way to explain why the microbiome behaves stably would be to argue that it does so as a consequence of the great number of species that compose it. The explanation in this case would work as follows: insofar as the microbiome is an ecosystem which is composed by a great number of species, and ecosystem biodiversity is believed to foster ecological stability under certain circumstances, then it will be expected that the human microbiome exhibits a stable behaviour (McCann 2000; Ives and Carpenter 2007). This way of accounting for the *explanandum* poses a serious challenge, though: the positive correlation between diversity and stability only works for non-random ecological communities; however, the opposite has been demonstrated to be true for random communities, in which an increase in biodiversity fosters *in*stability (May 1972). Because the microbiome is a random ecological community that due to its biological nature is expected to be suffering constant perturbations, then it will tend to be unstable. Therefore, what ecological theory predicts (instability) and what is empirically observed (stability) are at odds. The question that arises is then the following: what type of dynamics are instantiated in the microbiome so that its interactions result in a stable behaviour?

Fairly recently, Mougi and Kondoh (2012) have elaborated a model that overcomes the difficulty that May's results pose to explain the stability behaviour of random communities. In their view, the problem with May's model is that he only analysed communities with one interaction behavioural type (i.e. where all the members were

[4] In ecology, the concept of "stability" can be used to mean both that the number of species of the microbiome remains constant (i.e. that no species gets extinguished, also called *persistence*), and that the species density in the community recovers quickly after the community has been perturbed (i.e. once the density of one of the species in the community has slightly changed, also called *resilience*). A community whose species density remains constant is said to be in equilibrium. Obviously, if a community is stable in the second sense, it will also be stable in the first sense, but the opposite is not necessarily the case. In the case study that we present here, "stability" refers to the ability of the microbiome to recover its initial species density after a perturbation, i.e. it is a model to study resilience.

either mutualistic, or antagonistic, etc.), and they thought that different results might be obtained if the communities were studied taking into account the fact that there might be different behavioural types interacting simultaneously. They observed that, in fact, the existence of different combinations of interaction types in a community might be a solution to May's results, and thus allows showing that an increase in biodiversity *alone* (i.e. irrespectively of the interacting types) does not trigger instability. Mougi and Kondoh applied their reasoning to macroscopic communities, showing that in communities with different interaction types, an increase in biodiversity does not necessarily foster instability, if the proportion of cooperative types is high, thus explaining stability in terms of cooperation.

Coyte et. al's explanation of the stability of the microbiome follows the same logic as Mougi and Kondoh's research. They agree with them that a key element to explain the stability behaviour of the microbiome is the appeal to the existence of different interacting types. However, contrary to the claim by Mougi and Kondoh that cooperation fosters stability in macroscopic communities, Coyte et al. argue that it is *competition*, *not cooperation*, what explains the stability behaviour of microscopic communities. To prove their claim Coyte et al. develop a mathematical model and a series of computational simulations of the behaviour of the microbiome, and showed that the same result (i.e. that competition stabilizes and cooperation destabilizes communities) was observed irrespectively of the size of the community. Based on their analysis, they hypothesize that the destabilizing effect of cooperation is due to the strong dependencies among species that it generates, which would lead to the appearance of feedback loops in the community, whose destruction would lead the community to collapse. These feedback loops, on the contrary, would not appear when the species in the microbiome compete, thus making the community behaviour more stable. In this section, we analyse the mathematical model that Coyte et al. use to explain the stability of the microbiome.

Coyte et al. model the microbiome as a network of interconnected species to abstractly study its dynamics by applying a combination of linear stability analysis with computational simulations.[5] Their model describes the dynamics of density change of one species X_i given the interaction with another species X_j, and it is written as a modification of the Holling type 1 functional response:

$$\frac{dX_i}{dt} = X_i\left(r_i - s_i X_i + \sum_{j=1, j\neq i}^{S} a_{ij} X_j\right) \quad i = 1, \ldots, S \tag{1}$$

System of Eq. (1) expresses how the density of species i (X_i) will change over time given its own intrinsic growth rate (r_i) its interaction with other members of the same species, i.e. its self-interaction rate (s_i), which it is assumed to be the same for all species (i.e. $s_i = s$), and its interactions with the members of every other species,

[5] Their research consists in three different mathematical methods. In method 1 (linear stability analysis), they only consider communities that are close to equilibrium, while in methods 2 and 3 (permanence analysis, individual-based model) they investigate the behaviour of communities that are far from their equilibrium. Those two later methods yield the same results as the former (cooperation destabilizes communities). For reasons of space, we only consider method 1 for our analysis of the nature of explanation.

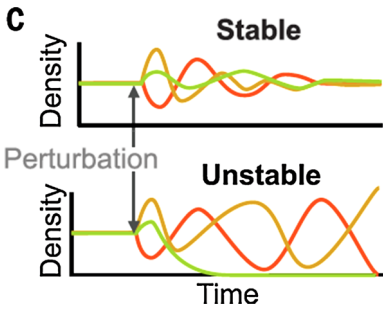

Fig. 2 Species-density variation due to a perturbation. Those communities that return to its equilibrium point are considered stable, and those that return to their previous equilibrium faster after a perturbation are considered more stable. Those that do not return to their equilibrium are deemed unstable. In this image there are two communities represented, stable and unstable. (From Coyte et al. (2015: Supplementary Figure 1C)

or interaction strength (a_{ij}, such that $j \neq i$). Finally, S expresses the total number of interacting species of a given community.

In Coyte et al.'s model, the interaction of a given species within the network will be determined by two parameters: first, the connectivity of i in the network, $C = [0, 1]$, defined as "the fraction of all S species that a single species i interacts with" (Coyte et al. 2015: Supplementary 4). Second, the nature of the interaction types between microbial species. They can take up to five possible forms, based on the signs of a_{ij}/a_{ji}: P_m (cooperation +/+), P_c(competition −/−), P_e (exploitation +/−), P^- (commensalism −/0) and P^+ (amensalism +/0). Accordingly, the proportion of interaction types between species must be established for a given community, being the total proportion of interaction types equal to one:

$$P_m + P_c + P_e + P^- + P^+ = 1 \tag{2}$$

Given this, Coyte et al. elaborate a phase portrait of their model to study its dynamical behaviour. To do so, they start by determining its equilibrium points, i.e. the points for which the variables of the system remain constant over time. Second, they determine the stability of each equilibrium point. An equilibrium point will be stable if after a small perturbation in the system the variables return to the same values that they had before the perturbation, and it will be unstable otherwise (Fig. 2). To determine the stability of each equilibrium point they perform a linear stability analysis. The linear stability analysis for an equilibrium point y is performed in three steps: (i) constructing the Jacobian matrix M of the system. The Jacobian matrix of a N dimensional system is a $N \times N$ square matrix whose elements m_{ij} will be given by the partial derivatives of the system. (ii) evaluating M at the equilibrium point $M|_y$ (iii) computing the eigenvalues of $M|_y$. Once the eigenvalues are computed, the stability of such equilibrium is determined by the following criterion: the equilibrium point y will be stable if and only if the real part of all the N eigenvalues of $M|_y$ is negative.[6]

When working with large dimensional systems the eigenvalues of the Jacobian matrix M tend to follow a concrete distribution. In the case of Coyte et al.'s model,

[6] We will use "stable points" to refer to what mathematically are defined as "asymptotically stable points".

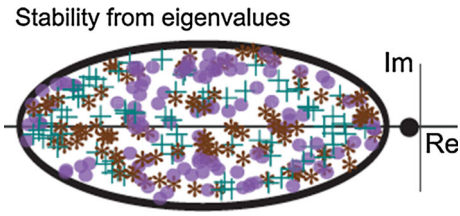

Fig. 3 Distribution of eigenvalues of a microbe community. The largest real part of the eigenvalues determines whether and how quickly the community will return to equilibrium after a perturbation. If this quantity is negative, the community is deemed stable. More negative implies more stability. The imaginary components determine the frequency of the oscillations in population densities after a perturbation. The different colors represent different simulations. (From Coyte et al. 2015: Supplementary Figure 1D)

the eigenvalues that they compute for each equilibrium fall into an ellipse of horizontal radius r_e in the complex plane and centre at $(-s, 0)$, being s the "average self-interaction", except for a single eigenvalue r_s which can lie outside (Fig. 3). Therefore, because an equilibrium requires all its eigenvalues to have a real negative part to be stable, an equilibrium in Coyte et al.'s model will be stable if and only if

$$\max(r_e, r_s) - s < 0 \tag{3}$$

Once we have briefly introduced a basis to study the stability of a given large dynamical system we can focus on analyse our case study. If we denote by y a given equilibria for Coyte's model, the Jacobian Matrix evaluated at that equilibria has the following entries (Coyte et al. 2015: Supplementary 3),

$$m_{ii} = -s \quad i = 1, \ldots, S \tag{4}$$

$$m_{ij} = a_{ij} \quad i, j = 1, \ldots, S$$

so that its eigenvalues—and therefore its stability—depend on the values of the self-interactions s and the a_{ij} terms, i.e. on the type of interactions between the species (cooperation, competition, etc.). Moreover, the connectivity also plays a role in the entries of the Jacobian matrix, so that the lower the connectivity of the network, the more a_{ij} terms will be equal to zero.

Coyte et al. are interested in computing the eigenvalues for each equilibrium point because the ecological stability behaviour of the system is characterized in terms of magnitudes directly related to the eigenvalues distribution of such equilibrium. First, if the equilibrium point is mathematically stable, then it will be ecologically stable too, i.e. the species density before the perturbation took place will eventually be recovered afterwards. In their context, a perturbation is produced when the densities of the species of community changes. Second, the behaviour of the community will be classified as more or less stable depending on how quickly it recovers its initial density distribution after the perturbation. Mathematically, this can be studied by analysing the eigenvalue distribution, such that the more negative the values of the distribution, the more attracting the equilibrium point will be, i.e. the faster the densities will go back to their initial states, and thus the system will be classified as more stable.

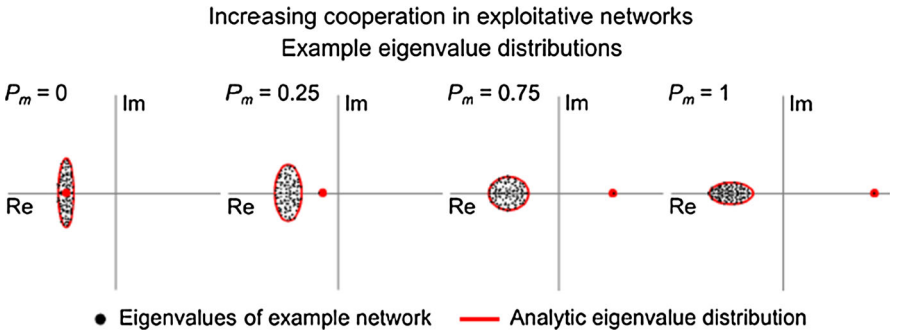

Fig. 4 Numerical confirmation of the analytical model for increasing cooperation in an exploitative community. The image shows how the distribution of the eigenvalues of a given exploitative community changes with respect to the increase in cooperative interactions. According to the linear stability analysis, a system will be more stable if the distribution of the eigenvalues lies further into the negative real part. Here, the single eigenvalue r_s shows how increasing the proportion of cooperative interactions destabilizes the system. (From Coyte et al. 2015: SM. Figure S3)

To study how the degree of ecological stability of each equilibrium point depends on the parameters r_i, s, a_{ij}, in system of Eq. (1), Coyte et al. run a series of simulations with different community types (exploitative, random and competitive). To do so, they define a measure of stability U, such that:

$$U = -\max(r_e - s, r_s) \tag{5}$$

Equation (5) gives the rightmost eigenvalue, i.e. the eigenvalue that represents the lower level of ecological stability. Right after, they study the behaviour of the following derivative $\frac{dU}{dP_m}$, that measures how U varies with respect to the proportion of cooperative interactions P_m. The sign of this derivative describes whether the community behaviour becomes more stable (if it is positive) or less stable (in case it is negative) in function of the proportion of interaction types of the species in the community. With this method, they show that for a given value of C and any given community type, gradually increasing cooperative interactions nearly always increases the overall return rate (the time it takes for the community to recover its initial densities after the perturbation), and the likelihood of the community being unstable (Fig. 4). This method also serves to prove the key role that C plays in the community behaviour: for a constant value of P_m, gradually increasing the value of C always has a destabilizing effect on the community. The overall result of their different simulations is that the higher the values of P_m and the higher the values of C, the less stable the community will be (Fig. 5).

A final step in Coyte et al.'s analysis is the ecological interpretation of their results, i.e. the explanation of what happens in the ecological community when the values of P_m and C increase so as to make the community less stable. In their view, communities with high values of C and P_m might generate strong dependencies among some of its members, which could be transformed into the existence of feedback loops in the community. The existence of feedback loops makes the community sensitive to small perturbations, insofar as a small change in the density of one of the species

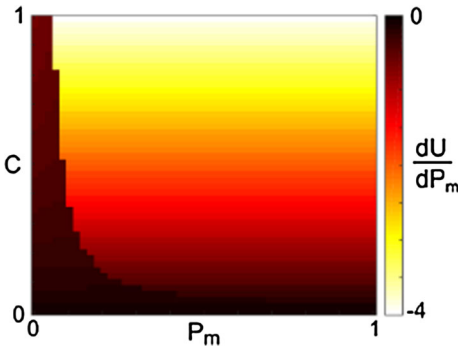

Fig. 5 Effect of increasing cooperation on stability in exploitative communities. This heat map shows how stability varies with respect to connectivity and the proportion of cooperation interactions in an exploitative community. The darker areas show higher stability, whereas the lighters show lower stability. It can be appreciated how stability is higher for the community with almost non-cooperative interactions. High connectivity makes the system unstable when cooperative interactions appear, for the combinations of both makes the system more vulnerable. The combination of low connectivity and high cooperation makes the system stable because of redundancy. (From Coyte et al. 2015: Supplementary Figure S4)

might trigger a cascade effect in the community that will in the end move it towards a different equilibrium point than the one it had before the perturbation. A low value of C and P_m avoids the creation of feedback loops and, therefore, has a stabilizing effect.

Coyte et al. rely on the empirical work done by Stein et al. (2013) to claim that their model is supported by empirical data. After the linear stability analysis and the different simulations they run, Coyte et al. make the following prediction: the proportion of destabilizing cooperative interactions in a stable microbiome has to be low in comparison with competitive and exploitative links, which will be predominant. The data presented by Stein et al. provides empirical validation for such hypothesis, thus suggesting that their model captures the right explanation of the stability of the microbiome.

4 Coyte's explanation as non-mechanist: explaining without providing a causal story

The case study presented above illustrates a type of explanation that we argue does not strictly follow the standard conception of scientific explanation as presented by the new-mechanists. To recapitulate, new-mechanists argue that to explain a phenomenon consist in: first, identifying a model of mechanism (individualized by its entities, its activities and their organization); second, identifying a causal story by means of which the model of mechanism produces the phenomenon to be explained. The question is now to determine in which sense (and to which extent) Coyte et al.'s explanation of the behaviour of the microbiome does not fulfil these two requirements. Particularly, we argue that even if a model of mechanism can be identified in Coyte et al.'s explanation, it is not the case that the explanatory force of the explanation they provide comes from a causal story being told.

Let us start by considering whether Coyte et al. individuate a model of mechanism. As we explained above, Coyte et al. model the microbiome as a network of interacting microbial species that behaves following the dynamics dictated by the system of Eq. (1). In their model, the entities are the different S microbial species that are part of the network, i.e. that compose the microbiome, and whose densities X_i are being studied. The activities of each of the entities are determined by the type of interactions they engage in (competition, cooperation, etc.), and would be given by the sign of a_{ij}. Finally, the organization of the community is given by the random network that describes the interactions among the species, including its number of nodes and its connectivity. Because the elements of the model of mechanism seem to be present in the case of the explanation given by Coyte et al., it seems that the first requirement to have a mechanistic explanation is satisfied by their model.

Once the model of mechanism has been specified it becomes necessary to show how the interactions among the different parts that compose it can produce the *explanandum*. In the case of Coyte et al. the *explanandum* is the stability behaviour of the microbiome, that is, how the species densities remain constant over time despite the existence of perturbations (i.e. some species that vary their densities). The *explanans* says that the key element that makes the microbiome stable is the existence of a high degree of competitive interactions among the species that compose it (so that competition *explains* stability). To get the appropriate connection between the *explanans* and the *explanandum* Coyte et al. proceed as follows: (i) they stipulate a community type (e.g. exploitative community); (ii) they determine its equilibrium points; (iii) they analyse the behaviour of the rightmost eigenvalue in function of the variation of the value of two key parameters of the topology [the connectivity C, and the proportion of interacting behaviours, as defined in (2)] in order to determine the degree of stability of each equilibrium point. In their analysis they observe that, for a fixed value of C, proportionally increasing P_m tends to make the system less stable (ecologically: less resilient). In other words, the time that the system will take to return to its initial equilibrium state after a perturbation will be larger, until a critical value of P_m is reached, such that the return time is equal to ∞, that is, the system becomes unstable.[7]

Now, the question for the new-mechanist is: does Coyte et al.'s mathematical model for the explanation of the stability behaviour of the microbiome actually captures *a causal story* of what happens in the microbiome so that it is ecologically stable? We suspect that the answer to this question is negative. First of all, because the way how Coyte et al. determine its stability behaviour is by means of a linear stability analysis of the system of ordinary differential equations (ODEs) specified in (1). And second, because even once the linear stability analysis is performed, what needs to be modelled to produce the *explanandum* is how the variations in P_m will affect the stability behaviour of the system.

Concerning the first point, the new-mechanist might argue that once the system is modelled, each ODE tells us a different causal story, so that the system as a whole is just an abstraction of the sum of all the individual causal stories modelled by each

[7] In their model, Coyte et al. do not exactly determine at which point the system will become unstable. It is enough for their explanation to show the general tendency of the community to an increasing value of *Pm*.

differential equation. But, even when all these causal stories are put together, they do not lead to the production of the phenomenon. At most, the ODEs give us information about how the variations in the densities of some species will respond to the variations in the densities of others. They do not give any information about how these variations are produced, nor how the variation of the densities of one species will affect the variation of the density of the rest of the species that compose the microbiome. By themselves, they say nothing about how the dependencies among the entities that compose the microbiome produce its stability. The only way in which the ODEs might say something about it is by numerically computing its evolution with respect to time (i.e. the trajectories of the system) when a minimal perturbation occurs.

However, for their explanation, Coyte et al. do not even consider the particular trajectories of the system. Their explanation only requires to study the stability of the equilibria by performing the mathematical steps described before (creating the Jacobian matrix, evaluating it, studying its eigenvalues, etc.), a procedure which does not require to specify the intermediate values of any of the variables X_i of the system. It is enough to study its long term behaviour, no matter which intermediate processes generate it. For this reason, it is difficult (if not impossible) to see how any of the steps followed by Coyte et al. describe a causal story that relates the specific interactions of the entity with the phenomenon to be produced.[8]

In relation to our second point, it seems to us that the explanation by Coyte et al. includes an additional step that makes it even harder to see how the causes can produce the stability behaviour. Concretely, once the ODEs generated by the system of Eq. (1) are solved, Coyte et al. still need to run simulations to determine how the variations in the proportions of the interacting species will affect the stability of the community. This step is crucial, because their *explanans* is precisely that a community will be stable if and only if it has the right proportion of interacting types. This step, as we said, is performed by analysing the variations of the rightmost eigenvalue to an increase in the proportion of P_m. But it is not specified which of the concrete species interactions will become cooperative, nor is that necessary to highlight the negative impact of cooperation on stability. The knowledge of the proportion of species that interact cooperatively is enough to establish their claim. Thus, again, the rehearsal of a causal story seems unnecessary to produce the phenomenon that Coyte et al. are explaining and therefore their explanation is not mechanistic.

5 Explaining with mathematics: combining topology with interaction types to explain stability behaviour

Coyte et al.'s model explains the stability behaviour of the microbiome but it does so in non-mechanistic terms. The question now is to determine how their model gains its explanatory force. In this section, we argue that what makes Coyte et al.'s model explanatory is the combination of the topological properties of the network instanti-

[8] Our argument in this section is inspired by a similar argument presented in Issad and Malaterre (2015: p. 284).

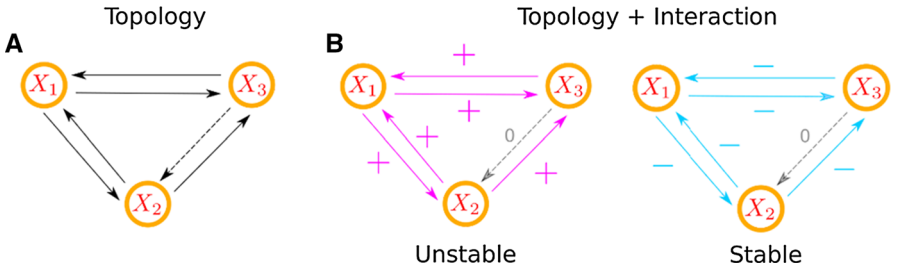

Fig. 6 Schematic representation of the microbiome network. Figure A represents the topology of the network by exclusively pinpointing the interactions among the species (the dashed line between X_3 and X_2 indicating that $a_{32} = 0$). Figure B represents the effect that considering not only the topology, but also the interaction types, might have on stability: for two communities with the same topology, only the one where competition is high (right) would be stable

ated by the microbiome with the knowledge about the dynamics that each of these topologies acquires in virtue of the combination of interacting types in the community.

Let us start by discussing the role of the topology in their *explanans*. In Coyte et al.'s model, the knowledge of the parameters s_i and a_{ij} allows determining the value of C and thus provides the topology of the system. Notice that if our network were non-random, i.e. if it realized a concrete topology, at this point we would have all the necessary information to decide whether the network is stable or unstable. Take the example of a scale-free network: were the microbiome a scale-free network, we could already identify its hubs (these nodes whose alteration would destabilize the network), as well as how it will behave if we increase the number of nodes, if we destroy some of the edges between nodes, etc. If this were the case, Coyte et al.'s explanation would be much simpler: the microbiome behaves stably *because* it realizes a scale-free network. This type of explanation would have the same nature as other topological explanations, gaining its explanatory force simply from the nature of the network that the system instantiates (Huneman 2010, 2018b; Jones 2014).

However, contrary to what happens in non-random networks, the microbiome instantiates a random network, which means that none of its topological properties will be known a priori. For that reason, Coyte et al. need to study the dynamics that the topology instantiates in order to determine which are the conditions that will make it stable. Furthermore, in their case, what ultimately explains whether the dynamics instantiated will be stable is the proportion of interacting types in the microbiome. That is, not every community with the same topology will be equally stable. Their explanation, therefore, needs to combine the determination of the topology of the system with the study of the dynamics that the variables that instantiate that topology will have under different conditions, i.e. for different interaction types (Fig. 6).

Therefore, Coyte et al.'s explanation of the stability behaviour of the microbiome would consists in something like:

> Ecosystem E instantiates a network N such that: (i) N corresponds to one of a set of topologies φ, and (ii) φ has the right proportion of interaction types. Thus, N has property P. Therefore, E also has P

Let us now develop what this model of explanation consists in. First, concerning (i), it is possible to mathematically compute which networks will make the microbiome stable, but this knowledge alone will render the *explanandum* unexplained, because there is still an element missing: which is the property in virtue of which N's topology is stable? Or, in other words, why N has a dynamics such that the system will tend to keep its stability? Appealing to its topology, by itself, will undermine the answer, because the system is purely random and there are multiple states that will make it stable, as well as multiple states that will make it unstable. But this option is not possible in the case of purely random networks, and therefore the explanatory force of Coyte et al.'s model, even if requiring the appeal to the topological structure of the microbiome, needs to be acquired from somewhere else.

Second, about (ii), our proposal is that the key additional element that makes Coyte et al.'s model fully explanatory of the stability behaviour of the microbiome is their appeal to the different interaction types that might appear in the network. The reason is that for the microbiome to be stable it is necessary that the *dynamics* instantiated by the topology are conducive to a stable equilibrium state. The only way of showing what are the conditions under which the dynamics instantiated by a random topology are conducive to a stable state is by studying their response to slight modifications in the parameters that define the system, namely a_{ij}. The way of computationally studying this is by evaluating the response(s) of the rightmost eigenvalue to changes in the values of a_{ij} in relation to changes in the proportions of interacting types in the community. In other words, it is necessary to study how the changes in the proportions of cooperative, competitive, exploitative, etc. interactions will affect the stability of the microbiome. Only once this response is studied, and the results are analysed, will the model be fully explanatory of the stability behaviour of the microbiome. In this sense, even if knowing the topology is necessary, since the dynamics of the system directly depends on the topology it instantiates, its knowledge is not sufficient to explain its behaviour. It is necessary to additionally understand how the community will respond to different proportions of interaction types. Only after this is done will the model gain its explanatory force, and Coyte et al. can assert that competition *explains* the stability behaviour of the microbiome.

Even if the points we have made here are relative to the case study we have analysed, we suspect that our conclusions about the nature of explanation can be extended to all the cases of behaviour explanations when they are given in terms of random networks. One key feature that distinguishes random from non-random networks is that the latter, but not the former, have specific properties derived from their network motifs and that can be unequivocally ascribed to every system that instantiates them. This does not mean however that random networks do not have network motifs: it means that their network motifs will highly depend on the way in which the network is constructed, and it will not be constant among all the systems that instantiate a random network. Because of this, we suspect that explanations of the behaviour of systems that instantiate a random network can only be given by studying an alternative mathematical property of the network (e.g. in Coyte et al.'s, the dynamics of the network), because: first, their network motifs will only be known once its mathematical properties have been studied; second, they will be highly variable depending on the mathematical properties that the network instantiates. This sharp contrast between random and non-random

networks is thus fundamental to understand why explanations in terms of the latter can be exclusively topological, whereas in the case of the former the topology has to be complemented with an appeal to dynamics.

6 A case for integrative pluralism in behaviour explanation

Until now we have argued that Coyte et al.'s explanation of the stability behaviour of the microbiome is neither purely mechanistic, since it lacks a causal story, nor exclusively topological, since the topology needs to be combined with a dynamics to be fully explanatory. We now argue that our case study supports an integrative pluralistic picture of behaviour explanation in biology. Following Mitchell (2003) and Brigandt (2010, 2013b; Brigandt et al. 2017), we take an explanation to be integrative when it requires the combination of concepts from different fields and of different types in order to gain its explanatory force. Such integration is normally driven by pragmatic considerations about the question asked: that is, as some scientific questions are about very complex phenomena, their responses usually require the integration of the knowledge of different fields to be fully satisfactory (Brigandt 2013a). In this sense, integrative pluralism is both beyond simple pluralism and against reductionism. Like simple pluralism, integrative pluralism accepts that some phenomena in ecology might be explained by simply telling a causal story (i.e. describing a mechanism), whereas others might be exclusively given in mathematical terms. However, integrative pluralism tries to go beyond this simple idea by showing how some complex phenomena are explained by integrating knowledge from different fields (ecology, population genetics, molecular biology, etc.), using both, causal mechanistic strategies and mathematical modelling. Against reductionism, integrative pluralism embraces the idea that even if the explanation of a complex phenomenon requires the appeal to knowledge from different fields, all of them are indispensable for the *explanans* to account for the *explanandum*. In other words, that the explanation of the phenomenon is only possible by integrating all the knowledge provided by the different fields, which would not be obtained if the knowledge of one of the fields were reduced to the knowledge of some of the others.

Following the tenets of integrative pluralism, we will make the point that the explanatory model of the stability behaviour of the microbiome presented by Coyte et al. exemplifies an integrative explanation. Concretely, their explanation combines a model of a mechanism with a mathematical model plus a series of computational simulations, integrating mathematical knowledge (network modelling, linear stability analysis), with knowledge about the patterns of interactions in ecological communities. We further argue that the necessity to integrate knowledge about the model of the mechanism with knowledge about the mathematical properties of the community (including its topology) is common for every scientific explanation that accounts for the phenomenon in terms of random networks.

As we argued in Sect. 4, Coyte et al.'s *explanans*, despite not telling any causal story of what happens in the microbiome, describes a model of the mechanisms enumerating which are the interacting entities, what type of activities they engage in, and the type of organization that the microbiome has. Also, as we argued in Sect. 5, their explanatory

model relies on the dynamics of the microbiome, which is acquired in virtue of the topology that it instantiates. The study of the behaviour of this dynamics determines a set of possible topologies, any of which will be stable, and thus one of the possible states the microbiome could be at. Importantly, both the model of the mechanism and the mathematical model are necessary and none of them is by itself sufficient without the other for Coyte et al.'s *explanans* to account for the *explanandum* in terms of "competition", as the authors say it does.

First of all, the model mechanism without the dynamics is completely unspecific about the stabilizing role of competition in the microbiome. Since they are explaining a behaviour (stability behaviour), and the behaviour is the result of the set of interactions between the entities (the species that compose the microbiome), it is necessary to study the way in which the interactions result in the behaviour that is observed. Or, in other words, detailing the entities that interact and the activities they engage in is insufficient to explain the stability behaviour if the way how these interactions make the system change is not studied adequately. For sure, the model of the mechanisms specifies that competition is one of the activities that the entities of the microbiome engage in and that can affect its stability. However, because the model of the mechanism in itself does not describe how the system changes in time, its specification is insufficient to make competition explanatory.

Secondly, the topology of the system, without the knowledge of the model of the mechanism, is also insufficient to explain why the microbiome exhibits a stable behaviour. As the community instantiates a *random* network, its network motifs have to be studied by analysing its dynamics. In the case of Coyte et al. they perform a linear stability analysis to study the reaction of the community to small perturbations. This analysis, however, does not make competition explanatory of the stability behaviour of the microbiome. As we explained extensively in Sect. 5, once Coyte et al. have determined the distribution of a_{ij} that make the microbiome stable, they have to study how the variations in the proportions of interaction types will affect the stability of the community. In other words, what makes competition explanatory in Coyte et al.'s model is not the strength of the interactions between the members of the microbiome, but the character (competitive, cooperative, exploitative, etc.) that those interactions have.

The previous observation entails that it is possible to have two communities with the same topology (i.e. with the same nodes, edges), but where only one of them is stable, whereas the other is not, due to the role that the interaction types have on the stability behaviour of the microbiome. Importantly, the knowledge of the influence of the interaction types on the dynamics that the topology instantiates is only possible once the model of the mechanism has been specified. Before this happens, and cooperation, competition, exploitation, etc. have been defined as possible activities of the entities, it is impossible to know whether the interaction types will have any influence on the dynamics of the network, since the latter is not the case for many other networks that might instantiate a random topology (e.g. the internet). Therefore, the knowledge of the dynamics is, by itself, insufficient to make competition explanatory, since the knowledge that competition will influence stability derives from the knowledge of the model of the mechanism of the microbiome.

From the two previous observations we derive that the explanation of the stability behaviour of the microbiome is a case of integrative pluralism, where a model of the mechanism needs to be combined with a mathematical analysis in order to be explanatory. As we argued in Sect. 5, the stability behaviour of the microbiome cannot be explained exclusively in terms of the topological properties of the network. Since it is a random network, it needs to be complemented with the study of the dynamics that the topology instantiates. As we argued there, this is a consequence of the fact that the network motifs of a random network can only be discovered a posteriori, by determining the effects that the interaction types have on the stability of the network. However, this claim about the study of the dynamics cannot be extended to every random network, since the reasons that make each random network stable will be different depending on their nature, and thus on the type of mathematical analysis that has to be done to determine its properties. The type of mathematical analysis will nevertheless depend on the model of the mechanism that is instantiated in each case, which will determine the nature of the entities that interact, as well as how their activities will be produced. The point we are making is thus that every scientific explanation that relies on the realization of a topology will be a case of integrative pluralism if the network that is realized is random.

7 Conclusion

The case study analysed in this paper fits well with the recent tendency in philosophy of science to emphasize the important role that mathematics play in some scientific explanations. Particularly, our case study, even if focused on behaviour explanations, shares many similarities with some of the cases of topological explanation analysed by Huneman (2010, 2018a, b). As in the cases he studies, the stability of the microbiome cannot be explained purely in mechanistic terms, due to the impossibility of telling a causal story that explains how the system behaves. Additionally, the explanation we analyse in this paper also gains its explanatory force from the specification of the topological properties of the system. However, and in contrast with the cases of explanations of the behaviour of non-random networks, in the case presented by Coyte et al. (which analyses the behaviour of a random network) the specification of the topology of the community is not enough to account for the *explanandum* (the stability behaviour of the microbiome). The authors are also required to study the dynamics that the topology instantiates, as well as to discover which set of topologies will make the system stable in virtue of its dynamics. Such set of topologies can only be discovered by studying the response of the rightmost eigenvalue to different proportions of P_m. Because the knowledge that the interaction types will influence the stability behaviour of the microbiome is only acquired after the model of the mechanism for the phenomenon is expelled out, Coyte et al.'s explanation constitutes a case of integrative pluralism. In other words, the explanatory force of their model is only gained from the combination of mathematical and mechanistic knowledge.

As a consequence, the analysis of our case study proves two main points: (a) the explanation of the behaviour of non-random ecological networks is different form the explanation of the behaviour of random networks. That is to say, because the network

is random, there is not any network type whose specification would automatically explain the behaviour of the system, and thus the network motifs have to be discovered; (b) behaviour explanations of random ecological networks are cases of integrative pluralism, in which knowledge from mathematics and from ecology are integrated to solve a complex problem. Concretely, the model of the mechanisms determines how to construct and study the mathematical model so that the *explanans* that accounts for the *explanandum* can be provided.

Acknowledgements A previous version of this paper was presented at the IX Conference of the Spanish Society for Logic, Methodology and Philosophy of Science (UNED, Madrid, 2018). We acknowledge all the participants for their helpful suggestions. José Díez, Johannes Findl, Christophe Malaterre, Eric Muszynski, and two anonymous reviewers provided insightful comments on a previous version of the paper. APC and JS are especially thankful to Sabina Leonelli, who had the idea of bringing together researchers from the Living Systems Institute and Egenis (University of Exeter), and thanks to whom the authors met and started to think about the issues that gave origin to this paper. The following institutions are formally acknowledged: RD and JS, Spanish Ministry of Economy and Competitiveness (FFI2016-76799-P); RD, Spanish Ministry of Economy and Competitiveness (BES-2013-063239); JS, Spanish Ministry of Education (FFU16/02570). APC, Spanish Ministry of Economy and Competitiveness (MTM2012-31714).

References

Alleva, K., Díez, J. A., & Federico, L. (2017). Models, theory structure and mechanisms in biochemistry: The case of allosterism. *Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences, 63,* 1–14.

Baker, A. (2005). Are there genuine mathematical explanations of physical phenomena? *Mind, 114,* 223–238.

Baker, A. (2009). Mathematical explanation in science. *The British Journal for the Philosophy of Science, 60,* 611–633.

Baker, A. (2015). Mathematical explanation in biology. In P. A. Braillard & C. Malaterre (Eds.), *Explanation in biology: An enquiry into the diversity of explanatory patterns in the life sciences* (pp. 229–247). Dordretcht: Springer.

Bechtel, W. (2010). The downs and ups of mechanistic research: Circadian rhythm research as an exemplar. *Erkenntnis, 73,* 313–328.

Bechtel, W. (2011). Mechanism and biological explanation. *Philosophy of Science, 78*(4), 533–557.

Bechtel, W., & Abrahamsen, A. (2005). Explanation: A mechanist alternative. *Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences, 36*(2), 421–441.

Bechtel, W., & Richardson, R. C. (1993). *Discovering complexity: Decomposition and localization as scientific research strategies*. Cambridge, MA: The MIT Press.

Brigandt, I. (2010). Beyond reduction and pluralism: Toward an epistemology of explanatory integration in biology. *Erkenntnis, 73*(3), 295–311.

Brigandt, I. (2013a). Explanation in biology: Reduction, pluralism, and explanatory aims. *Science and Education, 22*(1), 69–91.

Brigandt, J. (2013b). Systems biology and the integration of mechanistic explanation and mathematical explanation. *Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences, 44,* 477–492.

Brigandt, J. (2015). Evolutionary developmental biology and the limits of philosophical accounts of mechanistic explanation. In P. A. Braillard & C. Malaterre (Eds.), *Explanation in Biology: An Enquiry into the Diversity of Explanatory Patterns in the Life Sciences* (pp. 135–173). Dordretcht: Springer.

Brigandt, J., Green, S., & O'Malley, M. A. (2017). Systems biology and mechanistic explanation. In S. Glennan & P. Illari (Eds.), *The Routledge handbook of mechanisms and mechanical philosophy (Chapter 27)*. London: Routledge.

Coyte, K. Z., Schluter, J., & Foster, K. R. (2015). The ecology of the microbiome: Networks, competition, and stability. *Science, 350*(6261), 663–666.

Craver, C. F. (2007). *Explaining the brain*. New York: Clarendon Press.
Craver, C. F., & Darden, L. (2013). *In search for mechanisms: Discovery across the life sciences*. Chicago: University of Chicago Press.
Dethlefsen, L., & Relman, D. A. (2011). Incomplete recovery and individualized responses of the human distal gut microbiota to repeated antibiotic perturbation. *PNAS USA, 108*(1), 4454–4461.
Deulofeu, R., & Suárez, J. (2018). When mechanisms are not enough. The origin of eukaryotes and scientific explanation. In A. Christian, D. Hommen, N. Retzlaff, & G. Schurz (Eds.), *Philosophy of Science. European Studies in Philosophy of Science* (Vol. 9). Cham: Springer.
Díez, J. A. (2014). Scientific w-explanation as ampliative, specialised embedding: A neo-hempelian account. *Erkenntnis, 79,* 1413–1443.
Faith, J. J., Guruge, J. L., Charbonneau, M., Subramanian, S., Seedorf, H., Goodman, A. L., et al. (2013). The long-term stability of the human gut microbiota. *Science, 341*(6141), 1237439.
Foster, K. R., Schluter, J., Coyte, K. Z., & Rakoff-Nahoum, S. (2017). The evolution of the host microbiome as an ecosystem of a leash. *Nature, 548,* 43–51.
Glennan, S. (2002). Rethinking mechanistic explanation. *Philosophy of Science, 69*(S3), S342–S353.
Gonze, D., Coyte, K. Z., Lahti, L., & Faust, K. (2018). Microbial communities as dynamical systems. *Current Opinion in Microbiology, 44,* 41–49.
Human Microbiome Project Consortium. (2012). Structure, function and diversity of the health human microbiome. *Nature, 486*, 207–214.
Huneman, P. (2010). Topological explanations and robustness in biological sciences. *Synthese, 177,* 213–245.
Huneman, P. (2018a). Outlines of a theory of structural explanation. *Philosophical Studies, 175*(3), 665–702.
Huneman, P. (2018b). Diversifying the picture of explanations in biological sciences: Ways of combining topology with mechanisms. *Synthese, 195,* 115–146.
Huneman, P. (2018c). Realizability and the varieties of explanation. *Studies in History and Philosophy of Science*. https://doi.org/10.1016/j.shpsa.2018.01.004.
Issad, T., & Malaterre, C. (2015). Are dynamic mechanistic explanations still mechanistic? In P. A. Braillard & C. Malaterre (Eds.), *Explanation in biology: An enquiry into the diversity of explanatory patterns in the life sciences* (pp. 265–292). Dordretcht: Springer.
Ives, A. R., & Carpenter, S. R. (2007). Stability and diversity of ecosystems. *Science, 317*(5834), 58–62.
Jones, N. (2014). Bowtie structures, pathway diagrams, and topological explanations. *Erkenntnis, 79*(5), 1135–1155.
Kaplan, D. M., & Craver, C. F. (2011). The explanatory force of dynamical and mathematical models in neuroscience: A mechanistic perspective. *Philosophy of Science, 78*(4), 601–627.
Kuorikoski, J. (2007). Explaining with equilibria. In *Rethinking explanation* (pp. 149–162). Dordrecht: Springer.
Lange, M. (2013). What makes a scientific explanation distinctively mathematical? *The British Journal for the Philosophy of Science, 64,* 485–511.
Machamer, P., Darden, L., & Craver, C. F. (2000). Thinking about mechanisms. *Philosophy of Science, 67*(1), 1–25.
Marchesi, J. R., & Ravel, J. (2015). The vocabulary of the microbiome research: A proposal. *Microbiome, 3,* 31.
May, M. (1972). Will a large complex system be stable? *Nature, 238,* 413–414.
McCann, K. S. (2000). The diversity stability debate. *Nature, 405,* 228–233.
Mitchell, S. D. (2003). *Biological complexity and integrative pluralism*. Cambridge: Cambridge University Press.
Montoya, J. M., & Solé, R. V. (2002). Small world patterns in food webs. *Journal of Theoretical Biology, 214*(3), 405–412.
Mougi, A., & Kondoh, M. (2012). Diversity of interaction types and ecological community stability. *Science, 337*(6092), 349–351.
Nicholson, D. J. (2012). The concept of mechanism in biology. *Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences, 43*(1), 152–163.
Nikisianis, N., & Stamou, G. P. (2016). Harmony as ideology: Questioning the diversity-stability hypothesis. *Acta Biotheoretica, 64*(1), 33–64.
Potochnik, A. (2015). Causal patterns and adequate explanations. *Philosophical Studies, 172*(5), 1163–1182.
Rathkopf, C. (2018). Network representation and complex systems. *Synthese, 195,* 55–78.

Rice, C. (2012). Optimality explanations: A plea for an alternative approach. *Biology and Philosophy, 27*(5), 685–703.
Rice, C. (2015). Moving beyond causes: Optimality models and scientific explanation. *Noûs, 49*(3), 589–615.
Salmon, W. (1984). *Scientific explanation and the causal structure of the world*. Princeton: Princeton University Press.
Sober, E. (1983). Equilibrium explanation. *Philosophical Studies, 43*(2), 201–210.
Solé, R. V., & Montoya, J. M. (2001). Complexity and fragility in ecological networks. *Proceedings of the Royal Society B, 268*(1480), 2039–2045.
Sporns, O., Chialvo, D. R., Kaiser, M., & Hilgetag, C. C. (2004). Organization, development and function of complex brain networks. *Trends in Cognitive Sciences, 8*(9), 418–425.
Stein, R. R., Bucci, V., Toussaint, N. C., Buffie, C. G., Rätsch, G., Pamer, E. G., et al. (2013). Ecological modeling from time-series inference: Insight into dynamics and stability of intestinal microbiota. *PLoS Computational Biology, 9,* e1003388.
Suárez, J., & Deulofeu, R. (unpublished manuscript). Equilibrium explanation as structural non-mechanistic explanations: The case of long-term bacterial persistence in human hosts.
Walsh, D. A. (2015). Variance, invariance and statistical explanation. *Erkenntnis, 80*(3), 469–489.
Woodward, J. (2003). *Making things happen: A theory of causal explanation*. New York: Oxford University Press.
Woodward, J. (2017). Scientific explanation. In E. N. Zalta (Ed.), *The stanford encyclopaedia of philosophy*. https://plato.stanford.edu/archives/fall2017/entries/scientific-explanation/. Accessed 1 Nov 2018.
Xia, T. (2010). Network modelling in systems biology. Ph.D. Dissertation. https://lib.dr.iastate.edu/cgi/viewcontent.cgi?article=2611&context=etd. Accessed 16 Oct 2018.

Affiliations

Roger Deulofeu[1] · Javier Suárez[1,2] · Alberto Pérez-Cervera[3]

[1] Logos – Research Group in Analytic Philosophy, University of Barcelona, Barcelona, Spain

[2] Egenis – The Centre for the Study of Life Sciences, University of Exeter, Exeter, UK

[3] Department of Mathematics, Universitat Politècnica de Catalunya, Barcelona, Spain

teorema
Vol. XXXVIII/3, 2019, pp. 95-120
ISSN: 0210-1602
[BIBLID 0210-1602 (2019) 38:3; pp. 95-120]

Equilibrium Explanation as Structural Non-Mechanistic Explanations: The Case of Long-Term Bacterial Persistence in Human Hosts

Javier Suárez and Roger Deulofeu

RESUMEN

Philippe Huneman ha cuestionado recientemente los límites en la aplicación de los modelos mecanicistas de la explicación científica en base a la existencia de lo que denomina "explicaciones estructurales", en las que el fenómeno se explica en virtud de las propiedades matemáticas del sistema en que el fenómeno ocurre. Las explicaciones estructurales pueden darse en formas muy diversas: en virtud de la forma de *pajarita* (*bowtie*) de la estructura, de las propiedades topológicas del sistema, de los equilibrios alcanzados, etc. El papel que juegan las matemáticas en las explicaciones que apelan a la estructura de pajarita o a las propiedades topológicas del sistema ha sido recientemente examinado en varios trabajos. Sin embargo, el papel exacto que juegan las matemáticas en el caso de las explicaciones en términos de equilibrio aún no ha sido totalmente clarificado, y diferentes autores defienden interpretaciones contradictorias, algunas de las cuales las asemejarían más al modelo defendido por algunos filósofos mecanicistas que al modelo estructural de Huneman. En este trabajo, tratamos de cubrir ese déficit estudiando el papel que juegan las matemáticas en el modelo de equilibrio anidado (*nested equilibrium*) elaborado por Blaser y Kirchner para explicar la estabilidad de las asociaciones ontogenética y filogenéticamente persistentes entre humanos y microorganismos. De nuestro análisis se desprende que su modelo es explicativo porque i) se identifica una estructura matemática del sistema que viene dada por un conjunto de ecuaciones diferenciales que satisfacen una estrategia evolutivamente estable; ii) la estructura anidada del modelo hace que la estrategia evolutivamente estable sea robusta ante posibles perturbaciones; iii) esto es así porque las propiedades del sistema empírico son isomorfas a, pero no causalmente responsables de, las propiedades de la estrategia evolutivamente estable. La combinación de estas tres tesis hace que las explicaciones en términos de equilibrios se asemejen más al modelo estructural de explicación que al modelo mecanístico.

PALABRAS CLAVE: *explicación científica; mecanismos; explicación en términos de equilibrio; explicaciones estructurales; explicaciones no causales; estrategia evolutivamente estable*.

ABSTRACT

Philippe Huneman has recently questioned the widespread application of mechanistic models of scientific explanation based on the existence of structural explanations, i.e. explanations that account for the phenomenon to be explained in virtue of the mathematical properties of the system where the phenomenon obtains, rather than in terms of the mechanisms that causally produce the phenomenon. Structural explanations are very di-

verse, including cases like explanations in terms of bowtie structures, in terms of the topological properties of the system, or in terms of equilibrium. The role of mathematics in bowtie structured systems and in topologically constrained systems has recently been examined in different papers. However, the specific role that mathematical properties play in equilibrium explanations requires further examination, as different authors defend different interpretations, some of them closer to the new-mechanistic approach than to the structural model advocated by Huneman. In this paper, we cover this gap by investigating the explanatory role that mathematics play in Blaser and Kirschner's nested equilibrium model of the stability of persistent long-term human-microbe associations. We argue that their model is explanatory because: i) it provides a mathematical structure in the form of a set of differential equations that together satisfy an ESS; ii) that the nested nature of the ESSs makes the explanation of host-microbe persistent associations robust to any perturbation; iii) that this is so because the properties of the ESS directly mirror the properties of the biological system in a non-causal way. The combination of these three theses make equilibrium explanations look more similar to structural explanations than to causal-mechanistic explanation.

KEYWORDS: *Scientific Explanation; Mechanisms; Equilibrium Explanations; Structural Explanations; Non-Causal Explanations; Evolutionarily Stable strategy*.

In the last few years, a new trend in the debates about scientific explanation has flourished in philosophy of science. This new trend, "new-mechanism," emphasizes the role of mechanisms in scientific discourse in general, and in scientific explanation in particular [Machamer et al. (2000); Glennan & Illari (2017)]. Inspired by the developments in molecular biology, new-mechanists redefine causalism and argue that to explain a phenomenon consists in providing the mechanism that produces it. In the new-mechanist tradition, mechanisms are taken to be a set of *entities* (parts) and *activities* (operations) with a particular *organization* such that their causal interactions bring the phenomenon to be explained about [Glennan (2002); Bechtel & Abrahamsen (2005); Craver & Darden (2013); Craver (2007); Nicholson (2012); Issad & Malaterre (2015); Deulofeu & Suárez (2018)]. Thus, for a scientific explanation to be mechanistic, it must fulfill two necessary and sufficient conditions. First, it must identify a *model of mechanism* in which the mechanism is individuated by its parts, operations and organization. Second, it must provide a story of how the components of the mechanism are causally connected in such a way that they produce the *explanandum*.

New-mechanists share a basic commitment to a causal view of the world combined with: 1) the rejection of the Hempelian idea that explanations take the form of logical arguments, either inductive or deductive, and 2) the notion that mechanisms provide the causal "ingredient" that scientific explanations require to be genuinely explanatory[1]. Furthermore, they often assume a hierarchical view of mechanisms, acknowledging the existence of a diversity of scientific explanations in every science, thus

neither renouncing to the explanatory role of the special sciences, nor to the possible existence of mechanistic inter-level (hierarchical) explanations among different sciences [Krickel (2018)].

The wide scope of the New Mechanism account of scientific explanation in biology has been questioned due to the existence of explanations that seem to lack the causal ingredient that new-mechanists demand. One of the traditional explanatory types where this happens is in equilibrium explanations, where the mathematical properties of the empirical system (i.e. the fact that it reaches an equilibrium point) are taken as explanatory, irrespectively of the causal-mechanistic details of the system. Starting with Sober (1983), equilibrium explanations have been hypothesized to constitute an alternative to purely causal-mechanistic explanations [Batterman & Rice (2014); Rice (2015); Huneman (2018b), (2018c)]. However, it has also been argued that some equilibrium explanations admit a causal interpretation, if "causality" is understood in Woodward's interventionists terms [Woodward (2003); Kuorikoski (2007); Potochnik (2015)]. If the later were the case, as some new-mechanists are committed to an interventionist Woodwardian view of causation [Craver (2007); Kaplan & Craver (2011)], it could be argued: first, that the mathematical components that are present in equilibrium explanations describe the causal relationships among the entities of the system; second, that equilibrium explanations do not then constitute a real exception to the new-mechanist trend. The existence of these contradictory interpretations of the nature of equilibrium explanations (causal vs. non-causal) creates an important gap to understand how they gain their explanatory force, as well as about the specific role of causality in scientific explanation: is causality — at some level — a necessary ingredient in every scientific explanation, or are non-causal explanations also legitimate in certain cases?

In this paper, we aim to clarify this issue by studying Blaser & Kirschner's (2007) nested equilibrium model (NEM, hereafter) of the persistence of bacteria in human hosts. Our choice of this case is motivated by two reasons: on the one hand, Blaser & Kirschner's NEM explains the phenomenon in terms of the existence of an evolutionarily stable strategy (ESS, hereafter) among the different interacting organisms, a feature that makes it sufficiently analogous to most cases of equilibrium explanations reviewed in the philosophical literature so that our conclusion can shed light on the nature of scientific explanation; on the other hand, the explanatory force of their model is also conditional on the existence of a nestedness among different biological scales, i.e. on the

existence of a hierarchy of interrelated ESSs. As the acknowledgment of the existence of a hierarchy of mechanisms is a hallmark of the new-mechanist account of scientific explanation, and, to our knowledge, cases of nested equilibria have never been studied before in the philosophical literature, we believe that our case study could bring new light to the study of the old phenomenon of equilibrium explanations. Our aim is thus to analyse the explanatory role that the appeal to the existence of equilibria at different levels plays in the NEM. In that vein, we intend to provide a better understanding of the nature of equilibrium explanation, and to the role of causality in scientific explanation[2]. To do so, we frame the paper in the context of the debate between Huneman's structural account of scientific explanation and the causal-mechanistic account.

In section I, we introduce the general account of structural explanations presented by Huneman (2018a) and motivate the necessity of discussing the precise nature of equilibrium explanation to understand whether, and if so, to what extent, equilibrium explanations fit Huneman's account, or are rather a special case of causal-mechanistic explanations. In section II, we present our case study. In section III, we present our philosophical analysis. We first argue that the explanatory force of Blaser & Kirschner's NEM is mainly provided by the concept of ESS, plus the mathematical modelling that defines each strategy at each of the levels of the hierarchy, rather than by the causal-mechanistic details of the system. Additionally, the nested nature of the different ESSs plays a role in making the system robust to every possible intervention at different levels. Thirdly, and connected to this last point, we argue that no role is left for any causal element in their model, thus suggesting that their explanation constitutes a case of structural explanation as Huneman has defined it. Finally, in section IV, we present our conclusions.

I. Explaining With and Without Causes: The Role of Mathematics in Equilibrium Explanations

In recent years, the universal application of the "new-mechanist" account of scientific explanation in biology has been questioned on the basis of the existence of a family of explanations that do not rely on any causal features of the system whose properties they explain, but rather on its mathematical properties [Huneman (2010), (2018a), (2018b); Woodward (2013); Rice (2015); Kostic (2018), (2019); Deulofeu et al. (2019)]. Huneman has called these explanations "structural", and defines them as follows:

> Family of explanations for which the mathematical tools used in the description of an explanandum system belong to a mathematical structure whose properties are directly explanatory of some aspects of the system (such as equilibria, behaviour, limit regime, asymptotic behaviour, etc.) (…) They explain by accounting for the explananda through pinpointing structural relations that are mathematical relations of some sort. Mathematics here are not representing a dependence between structures in the world, but they are constituting the structural dependence itself, (…) and in virtue of that they are explanatory [Huneman (2018a), p. 695].

In contrast with mechanistic explanations, structural explanations do not include any mechanism, nor any causal story in their *explanans*. Furthermore, the inclusion of any of these elements would usually be taken as counterproductive to account for the *explanandum*. Structural explanations are abundant in systems biology, where an extensive amount of data has to be interpreted by using mathematical and computational tools [Green (2016), (2017); Green & Jones (2017); Brigandt et al. (2017)]. Huneman explicitly argues that some of the properties of the biological systems studied under the label of "systems biology" can only be explained by appealing to the formal (mathematical) properties that characterize those systems. A well-known example of this, studied by Jones (2014), is the vulnerability of the immunological system to attacks to the CD4+ T-cells. Drawing upon Kitano & Oda's (2006) case study, Jones argues that what explains the vulnerability of the human immune system to attacks on this particular component is its bowtie structure: because the human's immune system has a bowtie structure such that CD4+ T-cells are non-redundant elements in the core of the bowtie, the system is vulnerable to attacks on this type of cells (Figure 1). What is more important is that the vulnerability to attacks on CD4+ T-cells is not a consequence of the causal-mechanistic processes that produce the vulnerability: it is a consequence of the topological properties of the architecture (organization) of the immunological system. These topological properties determine its vulnerability to attacks on its core, as it is the only non-redundant element of the network, which is furthermore a necessary step for every other immunological process. Huneman summarized this kind of explanation as follows: "what is epistemically proper to this network modelling is that the topological properties found in the networks are such that they explain some of the properties one is interested in [vulnerability to attacks on CD4+ T-cells], (…) the instantiation of these properties is explained by the fact that the network is of such topological nature" [Huneman (2018b) p 127].

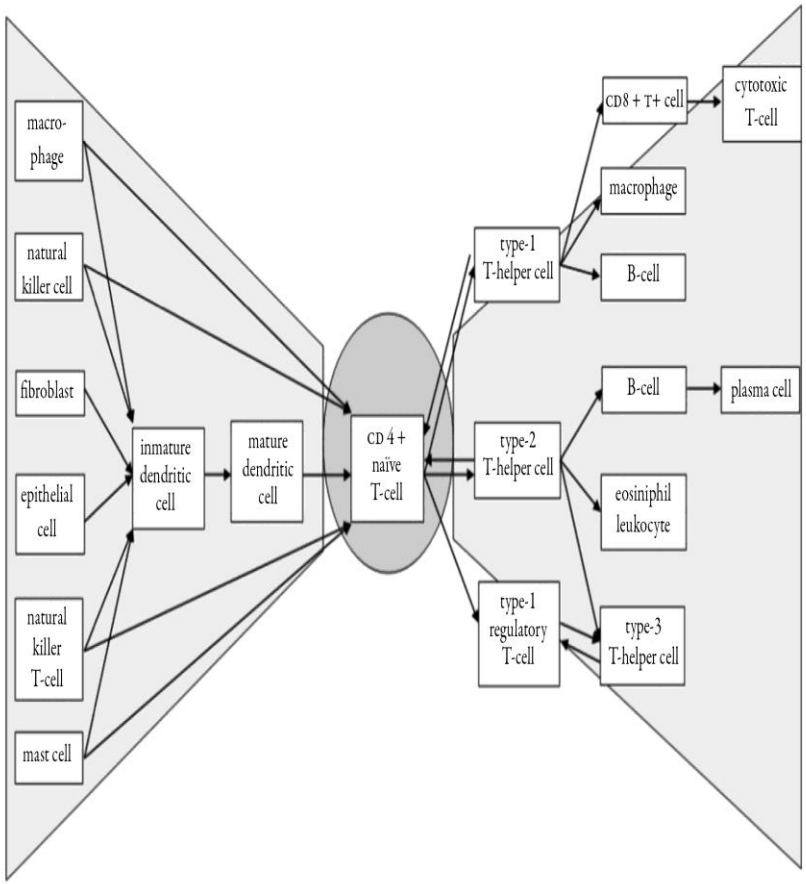

FIGURE 1. Bowtie structure of the immune system, with the CD4+ T-cells in the core of the bowtie. From Jones (2014), p. 1138, Fig. 1.

A second point that is epistemically proper to this kind of explanation is that the mechanisms that "sustain" the realization of such topological properties are irrelevant for explaining those properties (namely, the vulnerability of the network) [Huneman (2018c) pp. 6-8; Deulofeu et al. (2019); Moreno & Suárez, (submitted)][3].

Structural explanations are not restricted to cases of topological explanation, though. In his (2018c), p. 6, Huneman outlines the case of explanations in microeconomics, particularly the "ice cream vendors" problem — a direct application of the theory of Nash equilibrium to

human behaviour. In this situation, we imagine that there are two vendors standing on a beach and need to decide where to situate their stall in order to maximize their sales. Microeconomics, relying on game theory, says that the vendors will situate their stall in the middle of the beach, next to each other, to attract customers both in the area around them and in their extremes. By placing themselves in the middle of the beach, the vendors generate a Nash equilibrium, a situation where none of the players (the vendors) can change their strategy without decreasing their benefits (potential customers). Let us suppose we have to explain a scenario where there are two vendors placed in the middle of the beach. What explains the fact that both of them place their stalls in the middle? Huneman replies: "the fact that it simultaneously maximizes the share of each of them, or in other words, that it instantiates a Nash equilibrium." And adds: "[t]he mechanisms through which vendors move, decide, sell or buy, etc. are not explanatory relevant to this precise question" [Huneman (2018c), p. 6].

Nonetheless, Huneman just sketches the elements that make the Nash equilibrium explanatory in the case of the "ice cream vendors" but does not specify in detail what explaining with equilibria exactly entails, nor what is his reason to believe that mechanisms do not play any explanatory role in equilibrium explanations. Previous analyses of the role of equilibria in scientific explanations had been presented in Sober (1983) and Kuorikoski (2007). However, both authors reach opposing conclusions about where equilibrium explanations gain their explanatory force from: while the former argues that "equilibrium explanations show how the cause of an event can be (statistically) *ir*relevant to its explanation", and that their explanatory force comes exclusively from their mathematical structure [Sober (1983), p. 201], the latter believes that "explanations of singular events are indeed causal, even those supplied by equilibrium models" [Kuorikoski (2007), p. 149]. These opposing conclusions are interesting because they leave open whether equilibrium explanations must be considered a subtype of structural explanation (Sober), or a subtype of causal-mechanistic explanation (Kuorikoski), thus creating an important gap in how to understand the role of mathematics in this type of explanation. In addition to that, they leave open a question about the role of causality in scientific explanation in general for, if as Kuorikoski argues, even equilibrium explanations are in the end causal, then it could be argued that causality is a necessary ingredient in every genuine case of scientific explanation.

In the next section, we introduce Blaser & Kirschner's NEM of the persistence of bacteria in human hosts as a case study that we will use to motivate our response to these two questions.

II. A Nested Equilibrium Explanation of the Persistence of Bacteria in Human Hosts

Humans harbour an abundant number of microbes in their guts that constitute the human microbiome [Huttenhower et al. (2012); Lozupone et al. (2012)][4]. Among those microorganisms, some persist in our guts throughout our entire whole life cycle, whereas others are mainly transient, or appear in specific moments of our development, disappearing afterwards [Chiu & Gilbert (2015)]. Furthermore, some of those are hypothesized to have established long-term associations with humans over millions of years, with some people speculating that they might constitute co-evolved systems or hologenomes [Rosenberg & Zilber-Rosenberg (2014), (2016); Díaz (2015); Suárez (2018); Suárez & Triviño (2019); cf. Moran & Sloan (2015); Douglas & Werren (2016)]. Irrespectively of the evolutionary nature of those associations, the fact that organisms from different species engage in persistent long-term associations with each other is paradoxical from the perspective of the neo-Darwinian model of life and evolution. According to this model, when two individuals of different species associate, i.e. when they share the same habitat or niche, each one will pursue its own fitness interests. In this scenario, it might happen that the two organisms coexist peacefully for a period of time but, normally, peaceful coexistence will tend to break down: on the one hand, in the moment in which an opportunity for one of the organisms to benefit in detriment of the other appears, it will tend to grow to maximize its fitness until the other organism is destroyed (appearance of cheaters); on the other hand, it is also not infrequent that in a stable biological population where one out of two different survival strategies has been adopted among the members, the population becomes invaded by individuals that adopt an alternative strategy, until the point where the population collapses (external invasion). For these reasons, peaceful associations among organisms of different species are rare and will normally be short-term. Then, how is it possible that humans and some of their microbes establish persistent infections that are not disrupted by cheaters[5]? And which are the mechanisms that allow long-term associations that survive the challenges of sharing a habitat and are not perturbed by external invaders?

Blaser and Kirschner have recently developed a model "*to explain* the common features of microbial persistence in their human hosts" [(2007), p. 847, emphasis added)], i.e. to explain why humans and some specific microorganisms have overcome the difficulties of co-habitation[6]. They speculate that those situations represent a successful phenotype that must be maintained according to certain eco-evolutionary rules. In their view:

> persistence represents the evolved selection for balancing host and microbial interests, resulting in an equilibrium that, by definition, is long-term but not necessarily forever stable. We hypothesize that maintenance of this equilibrium requires a series of evolved, nested equilibria to achieve the overall homeostasis [Blaser & Kirschner (2007), p. 843].

They argue that such nested equilibria will be observed at different timescales: microscopic, at the level of the interactions between the immunological system of the host and cell-receptors of the microbes; mesoscopic, at the level of tissue function; tissue in which the microbe population inhabits; macroscopic, where evolutionary changes in the host and the microbe will occur to guarantee microbe transmission[7]. Blaser and Kirschner believe that any of these levels conforms to Nash equilibria in the form of an ESS that allows the persistence of the relationship. This is so because both the host and the microorganism will have developed a very specific hierarchy of cross-signalling mechanisms that generate a set of positive and negative feedback loops with each other that guarantee that the overall equilibrium is not disrupted.

Blaser and Kirschner's model begins by defining five populations at the microlevel whose changes with respect to certain variables are followed over time [see also Blaser & Kirschner (1999); Blaser & Atherton (2004); Blaser (2006)]. In the case of *Helicobacter pylori*, the variables include: *M*, which represents the population of mucus-living *H. pylori* (rate of change); *A*, which represents the *H. pylori* population that adhere to epithelial cells; *N*, which represents the concentration of nutrients available to bacteria derived from inflammation; *E*, which represents the concentration of effector molecules (molecules that the microbes generate to achieve some aims, such as suppressing immune response by the host); and *I*, that stands for the host response. Blaser and Kirschner's NEM includes five differential equations that track the changes in the variables of their model, as well as how they interact with each other[8].

For instance, to study how the concentration of mucus-living *H. pylori* varies over time due to the interaction with the other populations, they introduce the following differential equation:

$$\frac{dM}{dt} = g_m \alpha N(t) - \mu_m M(t) - \alpha M(t)(K - A(t)) + \delta A(t) \qquad (1)$$

where, g_m, α, μ_m and δ are parameters, whose value will depend on the situation; N, M, A (mentioned above) and K (the epithelial carrying capacity) are variables that together will determine the rate of change of the mucus-living population M. In (1), $g_m \alpha N(t) M(t)$ represents the potential growth of the population in virtue of the nutrient availability; $\mu_m M(t)$, represents the loss of *H. pylori* due to the process of mucus shedding; and $\alpha M(t)(K - A(t)) + \delta A(t)$ represents the potential loss/gain of *H. pylori* due to migration between the epithelial and the mucus-living populations. Obviously, migration from M to A can only happen when $A < K$, namely, when there is still room for more adherence to epithelial cells, and the opposite is the case for migration from A to M. Adherent sites are always limited or otherwise *H. pylori* would grow too much, risking the stability of the symbiotic association.

The inflammation induced by the bacteria on the host is captured by measuring the change of nutrient concentration over time:

$$\frac{dN}{dt} = \frac{b}{(b + I(t))} \beta E(t) - g_m N(t) M(t) - g_\alpha N(t) A(t) \qquad (2)$$

In (2), b, β, g_m and g_α are parameters. $N(t)$ is characterized by a gain term that is a function of the concentration of effector molecules, E, and the host response I. The equation shows the direct proportionality that exists between E and N, and the inverse proportionality between I and N. In other words, it shows the limiting effect that the host response has over the nutrient concentration, as well as the inducing effect of the bacteria on the nutrient concentration. (2) also specifies the rate of assimilation of nutrients of the mucus-living bacterial population and of the adherent epithelial populations.

Furthermore, for a microbe-host association to be *evolutionarily* persistent, the microbe needs to develop strategies for transmission. R_o captures this concept, quantifying "the transmission potential of a microparasite as the average number of secondary infections occurring when a single infectious host is introduced into a universally susceptible host population" [Blaser & Kirschner (2007) p. 844].

$$R_o = \frac{BN}{(x + b + v)} \tag{3}$$

In (3), *BN* measures the transmission rate as a function of the population size, *x* measures the rate of host mortality due to the microbe (measure of virulence), *b* is the rate of mortality of the host population independently of the microbe (measure of lifespan), and *v* is the rate at which the host recovers from the microbe infection (measure of immunity). Usually, for $R_o > 1$ microbial transmission is sustained whereas for $R_o < 1$ microbial transmission goes extinct.

Blaser and Kirschner show that in a persistent microbe-host association those five differential equations remain constant, and any deviation in one of the equations gets immediately counter-balanced by the adjustment of the other equations, keeping the equilibrium stable. Thus, Blaser and Kirschner claim this can only be possible if the system behaves according to a Nash equilibrium, and if the strategies followed by microbe and host conform to an ESS. Let us now see how an ESS can account *explanatorily* for observed constancy.

II.1. *The Role of the Evolutionarily Stable Strategy in Blaser and Kirschner's Model*

Nash equilibrium is a very common situation in game theory. It obtains when two players in a non-cooperative game adopt a strategy such that no individual change will render greater benefits to any of them, i.e. such that every change in the strategy that one of the players adopts independently will result in lower individual profit for that player. Nash equilibria are not necessarily, however, optimal strategies. It is sometimes possible to obtain a better net result if both players change their strategy simultaneously and a new equilibrium is reached. Nonetheless, this will only occur if *both partners* modify their strategy co-ordinately, but not if they do so independently. Therefore, no player has any incentive to modify his strategy individually. The prisoner's dilemma constitutes a typical example of a game whose solution is provided by a Nash equilibrium (Table 1). In this situation, two individuals — A and B — are accused independently of a crime, and each of them is interrogated separately and offered a deal: 1) if A betrays B and accuses her of having committed the crime, while B stays silent, A will have 4-years reduction of sentence and B will have no reduction (and the same, but inverted, occurs if B betrays A while A remains silent); 2) if both stay silent, each

of them will have a 3-years reduction of sentence; 3) if both betray each other, each will have a 1 years reduction of sentence. In this scenario, the Nash equilibrium is reached in situation 3), when both players betray each other. Of course, the result that they obtain is not optimal (each of them will only get 1 year reduction of sentence), but is such that none of them has any incentive to change her strategy individually, unless the other also does so, as otherwise she will have a bigger individual cost, i.e. she will have less years of reduced sentence [Nash (1950a), (1950b); Gintis (2000)].

A \ B	Betrays	Remains silent
Betrays	**1, 1**	4, 0
Remains silent	0, 4	**3, 3**

TABLE 1. Payoff matrix for the prisoner's dilemma. The numbers represent the amount of years that each subject would have as reduction of sentence. The optimal strategy is that where both remain silent (italics). Only the strategy where both betray constitutes Nash equilibria (bold).

An ESS is a biological strategy that, when it is adopted in a population, natural selection alone will keep the population safe from "intruder populations", in so far as the organisms that adopt an alternative strategy will be selected against. All ESSs are cases of Nash equilibria, but the opposite is not the case. If a solution to a non-cooperative game represents Nash equilibrium that is not an ESS, the solution could be disrupted by an alternative strategy that drives the population towards an alternative Nash equilibrium that constitutes an ESS [Smith & Price (1973); Smith (1974); Easley & Kleinberg (2010), pp. 209-227]. For instance, take the case of the stag hunt game (Table 2). This is a two players' game, where each player has two possible exclusive strategies: hunt-hares or hunt-stags. In this situation, there are three possible scenarios: 1) that both individuals are hare-hunters (case where both obtain a fitness benefit of 2); 2) that both individuals are stag-hunters (both obtain a fitness benefit of 3); 3) that one of the individuals is a hare-hunter whereas the other is a stag-hunter (in which case the hare-hunter obtain a fitness benefit of 3, whereas the stag-hunter obtains a fitness benefit of

0). In this situation, strategies 1) and 2) constitute a Nash equilibrium, for none of the players could get a better payoff by changing strategy. However, only 1) constitutes an ESS: while a hare-hunter and a stag-hunter do equally well when they are paired with a stag-hunter (fitness benefit of 3), hare-hunters score better than stag-hunters when they are paired with hare-hunters (hare-hunters score 2, while stag-hunters score 0). That means the stag-hunting strategy is not an ESS because if a hare-hunter is introduced in a population of stag-hunters, the population will evolve towards a population of hare-hunters. On the other hand, a population where all the individuals are hare-hunters represents an ESS, because if a stag-hunter is introduced in the population, it will be eventually extinct, for its fitness benefit will be lower than the fitness benefit of hare-hunters.

	Stag-hunter	Hare-hunter
Stag-hunter	**3, 3**	0, 3
Hare-hunter	3, 0	***2, 2***

TABLE 2. Payoff matrix for the stag hunt game. The numbers represent the net benefit for the individuals in the population that engage in the game. Cases where all the individuals in the population hunt exclusively stags or exclusively hares represent Nash equilibria (bold). However, only the case where both individuals hunt hares represent an ESS (italics).

Blaser and Kirschner apply this type of reasoning to persistent long-term host-microbe associations to argue that the situation must be the one that is obtained in Nash equilibrium, particularly in ESSs, where both positive and negative feedback between the host and the microbe occur, so that the equilibrium persists over time. The core idea of their model is that the equilibrium obtained at the microscopic level immediately affects the equilibrium at superior levels (mesoscopic and macroscopic). At the same time, the equilibrium at the higher levels affects in a specific way the possibility of new microbe-host persistent associations. The equilibria are nested and the association does not get in principle disrupted. The interaction among levels, partially captured by the equations (1)-(3), is as follows:

first, on the microscopic level one would find the microbial population, localized on an organ or tissue of the host, and the population of immune host cells responsible of recognizing the microbe population. The structure of both populations will depend on the nature of the original founder strain, the possibility for generating genetic variants, the selective pressures from other microbial cells in the same tissue and, more importantly, from the selection that the persistent microbe and the immune cells exert on each other [e.g. (Pradeu et. al 2013); Pradeu & Vivier (2016); Eberl (2016)]. The nature of the interactions between the organisms in the microscale will shape tissue function (or malfunction), and thus will partially determine the viability of the host, as well as the opportunity for microbial transmission (mesoscale). Finally, the effects of the microbe on the viability of the host will determine the host population structure (macroscale) that in return will affect microbial transmission (mesoscale) (Figure 2).

Even if the model illustrated in Figure 2 looks like a multilevel mechanism, for it appeals to a model of mechanism, it lacks the adequate type of causal stories that new-mechanists demand to have a proper explanation. First, because multilevel causation is mysterious, as Craver and Bechtel illustrate (2007), since causal relations happen exclusively intra-level. Second, because the type of inter-level readjustments of the system are symmetrical, occurring both top-down (e.g. from the macroscale to the mesoscale, or from the latter to the microscale), and bottom-up (e.g. from the microscale to the mesoscale, or from the latter to the macroscale), while relations between cause and effect are always asymmetrical. Third, because even if there could be a way to capture inter- and intra-level causal relations, this would be at odds with the information that NEM conveys and appeals to. NEM does not specify the causal way in which the entities at one level affect the entities at another level. It only specifies that the disruption of the equilibrium at one level will either prompt the collapse of the system (i.e. its death), or it will prompt the re adjustment of the equilibrium at that level due to the equilibria that exist in the other scales. In other words, NEM is not specific about how the equilibrium will be readjusted, it only predicts that it will be readjusted, provided that the other levels keep their equilibrium states. The causal elements (if any) that will bring this readjustment are irrelevant for the explanation of this behaviour in terms of NEM. What matters is exclusively the nested structure of the host-symbiont system (see section 4 for the full details).

In that vein, the nested structure of the model and the level of complex interactions between the different elements at the three scales (Nash equilibria, ESS) grant the persistence of the association. As it was

said before, one of the reasons why host-microbe associations do not normally last long is due to the presence of cheaters, organisms that enjoy the profits of the associations without paying the cost. Nash equilibria avoid the appearance of cheaters: cheaters are players that change their strategy unilaterally; in Nash equilibria, every player that does so is condemned to failure, and thus will be removed from the population. Furthermore, as the Nash equilibria that are reached in the population adopt the form of an ESS, it is not possible that an external invader adopting an alternative strategy disrupts the persistence of the association.

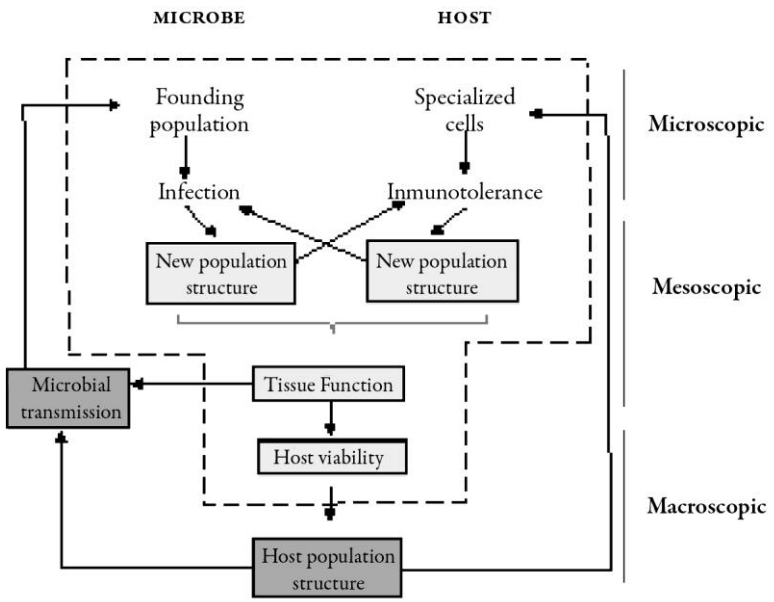

FIGURE 2. Nested equilibrium model. The dashed box represents those events that occur within the host. Adapted from Blaser & Kirschner (2007), p. 845, Fig. 2).

III. EQUILIBRIUM EXPLANATIONS AS STRUCTURAL AND NON-MECHANISTIC EXPLANATIONS

Blaser and Kirchner's NEM was developed to account for the persistence and the long-term character of certain human-microbe associations. Concretely, the authors seek to explain two paradoxes: first, why the association is not disrupted by the appearance of cheaters, i.e. entities that ben-

efit from the association without paying the costs; second, why the bacterial population is not entirely substituted by an intruder/external invader that deploys a different strategy. Only if those two phenomena are avoided, persistent host-bacterial associations can be successful. We will now argue that Blaser and Kirchner's NEM explains how those phenomena are avoided by appealing to mathematical, *but not causal*, properties, of host-microbial associations. In other words, we will argue that the alleged explanatory force of the NEM lies in the fact that: (i) it provides a mathematical structure in the form of a set of differential equations that together satisfy an ESS; (ii) that the nested nature of the ESSs makes the explanation of host-microbe persistence robust to any perturbation; (iii) that this is so because the properties of the ESS directly mirror the properties of the biological system in a non-causal way.

First of all, as shown in section II, Blaser and Kirschner's NEM consists in a series of differential equations that describe how the concentration of bacteria in different host tissues, their effector cells, their nutrient availability, the immunological response and their rate of transmission will change over time. These equations, as we explained, do not contain *a priori* any information about the persistence of the host microbe relationship. However, they provide information about how the different variables must be related to each other so that persistence obtains. Particularly, the equations measure the impact of host immunological response on bacterial colonization and, in doing so, allow determining the level at which host's response will abruptly disrupt colonization, as well as the levels at which bacterial inflammation will trigger a decrease in nutrient availability that in the end will disrupt colonization. And, in addition, they provide information about the way in which the solutions to these equations that guarantee the persistence of the symbiotic relation relate to: a) the rate of transmission of the symbiont (R_o), b) the viability of the host (tissue function and evolutionary advantages).

The set of equations can be resolved for a concrete host-symbiont system, and the evolution of the variables under study, as well as their interrelation, can be analysed. This will provide information about how they relate and how they are maintained constant, allowing predictions about empirical system[9]. However, notice that they would still provide no information about our *explanandum*, i.e. about what makes the host-microbe relationship persistent. To do so, the set of equations must be embedded in the framework of ESSs, i.e. it must model the biological situation as a non-cooperative game of two players, such that if any of the players (host, microbe) follows a unilateral strategy, the consequences

will be detrimental for the player that does so. That this is so can be seen by studying how changes in the equations that relate the concentration in nutrient availability, immune response, microbial concentration, etc. will relate to each other to make the system collapse if the change is unilateral. However, as we argued, the explanatory character of the equations comes exclusively from the possibility of embedding them in the framework of ESS. In other words, they are explanatory sound because it is possible to realize that no unilateral change that disrupts the system is possible without generating a chain reaction that either reverses the change or destroys the system. The ESS thus explains stability by ruling out two alternative scenarios: one where cheaters spread in the population, and another when an invader population entirely substitutes the actual one.

Second, the explanatory force of the ESS is reinforced in Blaser and Kirschner's NEM due to its nested nature. The nested nature of the equilibria works as a check and balances system which prevents that a disruption of the ESS at one of the levels (microscopic, mesoscopic and macroscopic) spreads across the other levels and destroys the host-microbe association. Let us explain this with an example: take the case of a disruption at the mesoscale that substitutes the microbe population for an invader. As we are at the mesoscale, the invader will disrupt tissue function in its own benefit, e.g. growing more than what the original microbial population would have grown, while at the same time escaping from the barriers of the immunological system. This type of change, totally beneficial for the bacteria at the mesoscale, would trigger two responses: First, a response at the macroscale that would be immediately detrimental for the bacteria. At this level, host viability, which is affected by the tissue function, will be reduced and, as a consequence, bacterial transmission will substantially decrease in relation to the transmission of those bacteria that cause no damage in tissue function. Secondly, at the microscale, where the invader population will not have generated immunotolerance, the invader population will be systematically blocked by the specialized immunological cells, especially the cells of the adaptive immune system. Furthermore, it is expected that the host will reduce nutrient availability, so that it affects in the long-run the intruders' population structure. Remember, as we said in section II, that the key of the ESS is that no player that changes its strategy unilaterally will be better. In this situation, even if the "player" might be better in one particular scale (mesoscale), the same will not be true for the other scales, and thus no possibility for invasion exists[10].

Third, and more concretely about the nature of ESS, we believe that Blaser and Kirschner's NEM, as any explanation that appeals to the existence of an ESS, explains the stability of host-microbe persistent associations in a non-causal way. Let us argue why we believe this to be so.

1) Blaser and Kirschner's NEM appeals to general properties of ESSs, and they make their model explanatory in virtue of the equivalence between the theoretical ESSs framework and the general properties of persistence host-symbiont associations. The strategy is the general strategy of Huneman's structural explanations: first, build a system *S'* whose properties match the properties of the real system *S* whose behaviour you aim to track. Second, study the behaviour of *S'* and attribute its properties to *S*. In Blaser and Kirschner's NEM, the strategy is applied as follows: first, build the ESS model for host-microbe persistent associations, as a case of a non-cooperative game of two players; second, study the behaviour of the ESS model, i.e. why the existence of an ESS, as the optimal solution for both players (Nash equilibrium), excludes the possibility of cheaters and invasive populations; third, attribute the properties of the ESS model to the empirical phenomenon, i.e. to empirical cases of host-microbe persistent associations. Notice that in this schema the explanatory force comes because the mathematical system that is built, in this case an equilibrium model, behaves in a certain way that (allegedly) is the way in which the empirical system will behave. But, importantly, it is irrelevant how the empirical phenomenon causally realizes the properties that it is attributed. And this is so in a double sense: on the one hand, because the NEM neither mention, nor needs to mention the specific species that interact to generate the ESS; on the other, because the causal connections between the entities (*if any*) are epistemologically irrelevant for the explanation of the phenomenon.

2) Despite the highly problematic way of identifying interlevel causal relations in a multilevel mechanism, as Craver and Bechtel (2007) explain, one could still try to appeal to Woodward's interventionist strategy to identify the supposed causes explaining the persistence of host-microbe associations. However, we believe NEM rules out the possibility of generating or even heuristically imagining any intervention *à la* Woodward, thus contradicting Kuorikoski and Potochnick's interpretation of equilibrium explanations. Let us explore this via an example. Recall that the *explanandum* is the phenomenon of persistence host-microbe associations. How would an intervention look like in Blaser and Kirchner's NEM? The only possibility would be to generate a situation such that the ESS disappears. However, no possible intervention is imaginable without destroying the system. Or,

in other words, any imaginable intervention that would make host-microbe associations non-persistent would directly change the system we are trying to explain, and thus the information it will provide will turn out to be irrelevant to account for the phenomenon. Recall the structure of ESS (Table 2). The only possibility of imagining a significant intervention would be via a change in the expected payoffs for the actions of each player. However, this intervention would not give any relevant information about why the association is stable in certain circumstance, because it would directly shift the focus of attention towards a new system, namely, one where there is not an ESS. Or, in other words, a causal explanation would consist in saying that the ESS is explanatory because if there were not an ESS the host-microbe association would not be stable. But this kind of reasoning is uninformative and, in our view, unexplanatory. The structural interpretation *à la Huneman*, on the contrary, offers a plausible account of how Blaser and Kirschner's NEM gains its explanatory force.

More importantly, the nested nature of the model, far from moving its explanatory force in a causal-mechanistic direction, generates the opposite effect. It just makes any possible intervention less imaginable. Because even if one causal intervention could be imagined for one specific level, how would it possibly work, if its effects would be cancelled out due to the existence of ESSs in the other levels? Or, in other words, how is it possible to imagine an intervention that causally escapes the inter-level connection? This connection is just a property of any host-microbe persistent association, and the explanatory power of the nestedness resides, precisely, in its possibility to cancel out the effect of every possible intervention. Therefore, we argue, a causal interpretation of the explanatory power of Blaser and Kirschner's NEM is not possible, since it would simply make the explanatory force of the model completely mysterious.

Of course, one might agree with what we just said, and still believe that our argument does not rule out the fact that the most appropriate interpretation of the explanatory force of Blaser and Kirchner's NEM is indeed causal. For instance, Blaser and Kirschner explicitly argue that specific host-microbe associations (human-*H. pylori*, human-*Salmonella typhi*, etc.) are "not necessarily forever stable" [(2007), p. 843], as obviously context (environment) matters, and in a changing context (environment) it is possible that concrete associations go selected against, simply because the environment selects against that coevolved system [see Díaz (2015); Suárez & Triviño (2019)]. In this context, it is possible to investigate the causes that made the system collapse, and if this is so, then the same must

be true for the cases in which the association is persistent. Nonetheless, we disagree, because that will entail changing the *explanandum* in two senses: first, making it specific to particular species; second, explaining the disruption of the persistence, instead of the persistence itself. And remember that our original *explanandum* was why some host-microbe associations are persistent, and the cases to rule out are the cases of cheaters and invasive populations. In our view, their model should be interpreted counterfactually: if a host-microbe association is persistent throughout the host's life cycle and evolutionarily long-term, then it will satisfy the conditions of the NEM reached through an ESS. And this situation will be so irrespectively of the species that interact, and thus irrespective of the causal-mechanisms that host and microbe could have developed to reach that equilibrium. As in the case of the ice vendors (section I), where the psychological mechanisms that have driven the vendors to put their stalls in the middle of the beach are explanatorily irrelevant to understand why their stalls are there, in the case of persistent associations causal-mechanistic details are simply superfluous. One can perfectly omit all those details and the explanation would still be epistemically sound.

Alternatively, an enumeration of the causes (if any) that would determine whether a concrete host-microbe association is stable will be irrelevant to explain its persistence if it is not conceived as a consequence of an ESS. This is because it would still be possible to imagine the existence of cheaters or invasive populations that deploy the same causal-mechanistic "machinery" to escape e.g. immunitary controls, without paying the cost of the symbiotic association. However, as we explained, because the host-microbe association constitutes a nested ESS, both the cheater and the invader population will end up disappearing from the population, just because the host-microbe persistent system has the structure that appears in the mathematical formulation of ESSs. Importantly, we are not here saying that Blaser and Kirchner's NEM rules out the possibility of telling a causal story of why concrete host-microbe associations are, sometimes, persistent, although some story about how to speak about interlevel causation should be provided.[11] Furthermore, we believe that such causal stories *could* be told to explain specific host-microbe associations, even when these must be complemented with the appeal to ESSs. Our point is rather *epistemological*: causal stories that seek to explain the existence of persistent host-microbe associations are neither required, nor explanatory in themselves. The element that provides the explanatory strength in equilibrium explanations is purely structural (in Huneman's terms), and it is connected with the possibility of accounting for the existence of an equilibrium (in Blaser and Kirschner's NEM, a nested ESS).

IV. Conclusion

In this paper, we have examined the explanatory force of equilibrium explanations, and have studied whether the explanatory force of equilibrium explanations can be better justified by applying the causal-mechanistic model of scientific explanation, or Huneman's structural model. Concretely, we have examined the role that mathematical vs. causal properties play in the explanation of the stability of persistent long-term host-microbe associations. Explaining the stability of this type of associations is paradoxical, as it requires explaining two facts: first, the absence of cheaters; second, the impossibility of the population being substituted by an intruder population. We have used Blaser and Kirschner's NEM to illustrate that the explanation of host-microbe persistent associations does not seem to be causal, but structural, relying solely on the non-causal mathematical properties of the association to explain its long-term persistence [Huneman (2018a), (2018b)]. We have argued that Blaser and Kirschner's NEM is explanatory of the long-term persistence of host-microbe associations because (i) it provides a mathematical structure in the form of a set of differential equations that together satisfy an ESS; (ii) that the nested nature of the ESSs makes the explanation of host-microbe persistence robust to any perturbation; (iii) that this is so because the properties of the ESS directly mirror the properties of the biological system in a non-causal way. In this vein, our case study shows how equilibrium explanations, even if nested, gain their explanatory force from the mathematical structure that describes the system, instead of from the causal interactions among its components. Our analysis supports two theses: first, that equilibrium explanations, even if nested (in a hierarchical setting), are structural rather than causal-mechanistic; second, that causality, even if necessary in some explanations, is not a universally necessary requirement of every scientific explanation.

Logos – Research Group in Analytic Philosophy / Barcelona Institute of Analytic Philosophy
Department of Philosophy
University of Barcelona
C/Montalegre nº6, 08001. Barcelona, Spain
E-mail: javier.suarez@ub.edu
E-mail: roger.deulofeu@gmail.com

ACKNOWLEDGMENTS

A previous version of this paper was presented in the meeting 'Process epistemology: A workshop with Bill Bechtel', University of Exeter, May 2017. The authors want to thank all the participants for their feedback, and especially to Bill Bechtel for a wonderful discussion. José Díez, John Dupré, Philippe Huneman, Thomas Pradeu, Johannes Findl and two anonymous reviewers are acknowledged for their comments on earlier versions of the paper. The following institutions are formally acknowledged: Javier Suárez and Roger Deulofeu, Spanish Ministry of Economy and Competitiveness (FFI2016-76799-P); Roger Deulofeu, Spanish Ministry of Economy and Competitiveness (BES-2013-063239); Javier Suárez, Spanish Ministry of Education (FFU16/02570).

NOTES

[1] The commitment to a causal view of the world does not entail either a physical reductionism [as in Salmon (1984)] or an "ontic" interpretation of scientific explanation [as in Craver (2014)]. Cf. Glennan (2002), Bechtel & Abrahamsen (2005), for a model-based interpretation of mechanisms.

[2] There are other cases where equilibrium models have been used to explain the stability of biological associations [Baalen & Jansen (2001); Selosse *et al.* (2006)]. We have chosen to analyse Blaser & Kischner's NEM for its generality, and because it is a case of equilibrium explanation generally accepted among biologists. Nonetheless, our conclusions also apply to these cases. Thanks to Philippe Huneman for pointing this fact to us.

[3] Following Brigandt (2013), we consider that an element of an *explanans* is explanatory relevant if and only if removing it from the explanation entails that the *explanandum* does not follow, and it's explanatory irrelevant otherwise [(2013), p. 480].

[4] "Microbiota" refers to "[t]he assemblage of microorganisms present in a defined environment", and "microbiome" is used to denote "the entire habitat, including the microorganisms (bacteria, archaea, lower and higher eukaryotes, and viruses), their genomes (i.e., genes), and the surrounding environmental conditions" in a given environment [Marchesi & Ravel (2015), p. 1]. For the purposes of this paper, we will not distinguish the two concepts, and they will be used to refer only to the community of microorganisms present in a given environment.

[5] In biology, persistent infection refers to lifelong associations between a host and some species of microbes that do not necessarily harm the host, although they might do it in the long-term. The term should not be confused with its medical use, where "infection" is usually employed in reference to pathogens, or disease-causative agents.

[6] Their model is in principle developed exclusively for pair associations, between one host and one microorganism.

[7] Those different levels have both a temporal and a scale correlation: the macroscale refers to the evolutionary time, the mesoscale refers to organismal development and the microscale refers to the interactions among different cell types.

[8] Since our purpose is only to illustrate the main features of the model and their relation to Blaser and Kirschner's explanation, for a matter of simplicity we only introduce two of the equations.

[9] Information about the values that the variables must take for a concrete (empirically real) host-microbe association, if the association is known to be stable.

[10] It exists, but if and only if the intruder changes the situation *in the three scales*. That is precisely the nature of the nested model.

[11] See Craver & Bechtel (2007) for a proposal.

REFERENCES

BATTERMAN, R.W., and C. C. RICE (2014), "Minimal Model Explanations"; *Philosophy of Science* 81. 3, pp. 349-376.

BECHTEL, W., and A. ABRAHAMSEN (2005), "Explanation: A Mechanist Alternative"; *Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences* 36 (2), pp. 421–441.

BLASER, M. J. (2006), "Who Are We? Indigenous Microbes and the Ecology of Human Diseases"; *EMBO Rep* 7, pp. 956–960.

BLASER, M. J. and J. ATHERTON (2004), "*Helicobacter pylori* Persistence: Biology and Disease"; *J. Clin. Invest*. 113, pp. 321–333.

BLASER M. J., and D. KIRSCHNER (1999), "Dynamics of *Helicobacter pylori* Colonization in Relation to the Host Response"; *Proc Nat Acad Sci* 96, pp- 8359-8364.

— (2007), "The Equilibria that Allow Bacterial Persistence in Human Hosts"; *Nature* 449, pp. 843–849.

BRIGANDT, J. (2013), "Systems Biology and the Integration of Mechanistic Explanation and Mathematical Explanation"; *Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences* 44, pp. 477–492.

BRIGANDT, J., S. GREEN, and M. A. O'MALLEY (2017)," Systems Biology and Mechanistic Explanation"; in S. Glennan and P. Illari (eds.) *The Routledge Handbook of Mechanisms and Mechanical Philosophy*, London: Routledge (chapter 27).

CHIU, L., and S. F. GILBERT (2015), "The Birth of the Holobiont: Multi-species Birthing Through Mutual Scaffolding and Niche Construction"; *Biosemiotics* 8 (2), pp. 191–210.

CRAVER, C. F. (2007), *Explaining the Brain*; New York: Clarendon Press.

— (2014), "The Ontic Account of Scientific Explanation"; in Kaiser, M. I., Scholz, Plenge, R. D. Hüttemann, A. (eds.), *Explanation in the Special Sciences: The Case of Biology and History*, Springer Verlag. pp. 27–52.

CRAVER, C. F., and W. BECHTEL (2007.), "Top-Down Causation Without Top-Down Causes"; *Biology & Philosophy* 22(4), pp. 547-563.

CRAVER, C. F., and L. DARDEN (2013), *In search for Mechanisms: Discovery Across the Life sciences*; Chicago: University of Chicago Press.

DEULOFEU, R. and J. SUÁREZ (2018), "When Mechanisms Are Not Enough: The Origin of Eukaryotes and Scientific Explanation"; in Christian A., Hommen D., Retzlaff N., Schurz G. (eds) *Philosophy of Science*. European Studies in Philosophy of Science, vol 9. Springer, Cham.

DEULOFEU, R, J. SUÁREZ and A. PÉREZ-CERVERA (2019), "Explaining the Behaviour of Random Ecological Networks: The Stability of the Microbiome as a Case of Integrative Pluralism"; *Synthese. https://doi.org/10.1007/s11229-019-02187-9.*

DÍAZ, J. S. (2015), "El Mecanismo Evolutivo de Margulis y los Niveles de Selección"; *Contrastes* XX (1), pp. 7-26.

DOUGLAS A. E. and J. H. WERREN (2016), "Holes in the Hologenome: Why Host-Microbe Symbioses Are Not Holobionts"; *mBio* 7 (2), e02099-15.

EASLEY, D. and KLEINBERG, J. (2010), *Networks, Crowds, and Markets: Reasoning about a Highly Connected World*; Cambridge University Press.

EBERL, G. (2016), "Immunity by Equilibrium"; *Nat. Rev. Immunol.* 16, pp. 524-532.

GINTIS, H. (2000), *Game Theory Evolving: A Problem-Centered Introduction to Modeling Strategic Behavior*.; Princeton University Press.

GLENNAN, S. (2002), Rethinking Mechanistic Explanation; *Philosophy of Science* 69 (S3), pp. S342–S353.

GLENNAN, S. and ILLARI, P. (Eds.) (2017), *The Routledge Handbook of Mechanisms and Mechanical Philosophy*, Taylor and Francis.

GREEN, S. (2016), *Philosophy of System Biology*; Dordrecht: Springer.

— (2017), "Philosophy of Systems and Synthetic Biology"; *The Stanford Encyclopedia of Philosophy* (Edition Spring 2019), Edward N. Zalta (ed.), URL = <*https://plato.stanford.edu/archives/spr2019/entries/systems-synthetic-biology/*>.

GREEN, S. and JONES, N. (2016), "Constraint-Based Reasoning for Search and Explanation Strategies for Understanding Variation and Patterns in biology"; *Dialectica* (70)3, pp. 343-374.

HUNEMAN, P. (2010), "Topological Explanations and Robustness in Biological Sciences"; *Synthese* 177, pp. 213–245.

— (2018a), "Outlines of a Theory of Structural Explanation"; *Philosophical Studies* 175 (3), pp. 665–702.

— (2018b), "Diversifying the Picture of Explanations in Biological Sciences: Ways of Combining Topology with Mechanisms"; *Synthese* 195, pp. 115–146.

— (2018c); "Realizability and the Varieties of Explanation"; *Studies in History and Philosophy of Science. <https://doi.org/10.1016/j.shpsa.2018.01.004>*

HUTTENHOWER C, GEVERS D, KNIGHT R, CREAS HH, *et al.* (2012) "Structure, Function and Diversity of the Healthy Human Microbiome"; *Nature* 486, pp. 207–214.

ISSAD, T., and C. MALATERRE (2015.), "Are Dynamic Mechanistic Explanations Still Mechanistic?"; in P. A. Braillard and C. Malaterre (eds.) *Explanation in*

Biology: An Enquiry into the Diversity of Explanatory Patterns in the Life Sciences. Dordrecht: Springer, pp. 265–292.

JONES, N. (2014), "Bowtie Structures, Pathway Diagrams, and Topological Explanations"; *Erkenntnis* 79 (5), pp. 1135–1155.

KAPLAN, D. M. and C. F. CRAVER (2011), "The Explanatory Force of Dynamical and Mathematical Models in Neuroscience. A Mechanistic Perspective"; *Philosophy of Science* 78.4, pp. 601-627.

KITANO, H., and K. ODA (2006), "Robustness Trade-Offs and Host-Microbial Symbiosis in the Immune System"; *Molecular Systems Biology* 2, pp. 1–10.

KOSTIĆ, D. (2018), "The Topological Realization"; *Synthese*, 195(1), pp. 79-98.

— (2019), "Minimal Structure Explanations, Scientific Understanding and Explanatory Depth", *Perspectives on Science*, 27 (1), pp. 48-67.

KRICKEL, B. (2018), *The Mechanical World: The Metaphysical Commitments of the New Mechanistic Approach*; (Vol. 13). Springer.

KUORIKOSKI, J. (2007), "Explaining with Equilibria"; in Persson, J., and Ylikoski, P. (Eds.), *Rethinking explanation*; Springer, Dordrecht, pp. 149-162.

LOZUPONE, C. J. I. STOMBAUGH, J. I. GORDON, J. K. JANSSON, and R. KNIGHT (2012), "Diversity, Stability and Resilience of The Human Gut Microbiota"; *Nature* 489 (7415), pp. 220–230.

MACHAMER, P., DARDEN, L., and C.F. CRAVER (2000), "Thinking About Mechanisms"; *Philosophy of science*, 67(1), pp. 1-25.

MARCHESI, J. R., and J. RAVEL. (2015), "The Vocabulary of the Microbiome Research: A Proposal"; *Microbiome* 3, p. 31.

MORAN, N., and D. B. SLOAN (2015), "The Hologenome Concept: Helpful or Hollow?"; *PLoSBiol* 13 (12), e1002311.

MORENO A., and J. SUÁREZ (submitted), "Plurality of Explanatory Strategies in Biology: Mechanisms and Networks".

NASH, J. F. (1950a), "The Bargaining Problem"; *Econometrica: Journal of the Econometric Society* 18 (2), pp. 155-162.

— (1950b), "Equilibrium Points in N-Person Games"; *Proceedings of the National Academy of Sciences* 36 (1), pp. 48-49.

NICHOLSON, D.J. (2012), "The Concept of Mechanism in Biology"; *Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences* 43 (1), pp. 152–163.

POTOCHNIK, A. (2015) "Causal Patterns and Adequate Explanations"; *Philosophical Studies*, 172(5), pp. 1163-1182.

PRADEU, T., JAEGER, S., and E. VIVIER (2013), "The Speed of Change: Towards a Discontinuity Theory of Immunity?"; *Nature Reviews Immunology*, 13(10), p. 764.

PRADEU, T., and E. VIVIER (2016), "The Discontinuity Theory of Immunity"; *Sci. Immunol.* 1 (1): aag0479.

RICE, C. (2015), "Moving Beyond Causes: Optimality Models and Scientific Explanation"; *Noûs* 49.3 pp. 589-615.

ROSENBERG E and I ZILBER-ROSENBERG (2014), *The Hologenome Concept*. London, Springer.

— (2016) "Microbes Drive Evolution of Animals and Plants: The Hologenome Concept"; *mBio* 7 (2): e01395-15.

SALMON W. (1984), *Scientific Explanation and the Causal Structure of The World*; Princeton: Princeton University Press.

SELOSSE, M. A., RICHARD, F., HE, X., and S.W. SIMARD (2006), "Mycorrhizal Networks: des Liaisons Dangereuses?"; *Trends in Ecology and Evolution* 21 (11), pp. 621-628.

SMITH, J. M. (1974), "The Theory of Games and the Evolution of Animal Conflicts"; *Journal of Theoretical Biology*, 47(1), pp. 209-221.

SMITH, J. M., and G. R. PRICE (1973), "The Logic of Animal Conflict"; *Nature*, 246(5427), p. 15.

SOBER, E. (1983), "Equilibrium Explanation"; *Philosophical Studies* 43.2, pp. 201-210.

SUÁREZ, J. (2018). "The Importance of Symbiosis in Philosophy of Biology: An Analysis of the Current Debate on Biological Individuality and its Historical Roots"; *Symbiosis* 76(2) pp. 77-96.

SUÁREZ, J. and V. TRIVIÑO (2019), "A Metaphysical Approach to Holobiont individuality: Holobionts as Emergent Individuals"; *Quaderns de Filosofia* 6(1), pp. 59-76.

VAN BAALEN, M. and V.A. JANSEN (2001), "Dangerous Liaisons: The Ecology of Private Interest and Common Good": *Oikos*, 95(2), 211-224.

WOODWARD, J. (2003), *Making Things Happen: A Theory of Causal Explanation*; New York: Oxford University Press.

— (2013), "Mechanistic Explanation: Its Scope and Limits"; *Aristotelian Society Supplementary Volume* 87 (1), pp. 39–65.